# Estimating Sparse Representations from Dictionaries With Uncertainty

Vom Fachbereich 18
Elektrotechnik und Informationstechnik
der Technischen Universität Darmstadt
zur Erlangung der Würde eines
Doktor-Ingenieurs (Dr.-Ing.)
genehmigte Dissertation

von
Christian Weiß, M.Sc.
geboren am 21.06.1985 in Wiesbaden, Deutschland

| | |
|---|---|
| Referent: | Prof. Dr.-Ing. Abdelhak M. Zoubir |
| Korreferent: | Prof. Dr. Fulvio Gini |
| Tag der Einreichung: | 06.12.2016 |
| Tag der mündlichen Prüfung: | 24.03.2017 |

D 17
Darmstadt, 2017

# Acknowledgments

I would like to thank all the people who accompanied and supported me during my doctoral study in various ways on a professional and private level.

First, I thank my supervisor, Prof. Dr.-Ing. Abdelhak M. Zoubir, for his trust, for the time and freedom of pursuing my research interests, and for the opportunity of accomplishing my Ph.D. at the Signal Processing Group. I am very grateful for the chance to extend my international experiences by visiting conferences and by spending time in Taiwan for a research visit during the Ph.D program. Hereby, I also thank the Graduate School of Computational Engineering at Technische Universität Darmstadt for the scientific and financial support. In addition, I wish to thank Prof. Dr. Chong-Yung Chi and the members of the Wireless Communications and Signal Processing Laboratory at the National Tsing Hua University, Hsinchu,Taiwan, for their great hospitality during my research visit. It was truly a pleasure.

A special thanks goes to my co-supervisor, Prof. Dr. Fulvio Gini, for offering valuable feedback. I also thank Prof. Dr. Sebastian Schöps and Prof. Dr.-Ing. Ralf Steinmetz for their interest in my work and for being in the Ph.D. committee.

Further, I would like to express my sincere gratitude to Prof. Dr. Shinji Yamashita from the Yamashita laboratory of photonic communication devices at the University of Tokyo, Japan, for providing valuable sensor data that I could use for my research. I also thank Dr. Kevin Brinkmann, Dr. Sebastian Knabe, and Dr. Florian Schulz for the industrial cooperation along with fruitful discussions and support, which have led to interesting results. The accomplishment of the related industry projects would not have been possible without the help of Vineet Kumar, Sebastian Stiefel, and Mark Ryan Balthasar.

A great thanks also goes to all current and former members of the Signal Processing Group at Technische Universität Darmstadt. In particular, I would like to thank Jürgen Hahn, Adrian Šošić, Di Jin, Nevine Demitri and Dominik Reinhard for all the interesting, funny and fruitful discussions in the office. I will not forget our supper meetings at *El Cid*, the special lunch hours with Simon Rosenkranz, and the long working nights with Feng Yin and Fiky Suratman. These are truly memorable experiences.

I also enjoyed working with Sara Al-Sayed, Mouhammad Alhumaidi, Mark Ryan Balthasar, Patricia Binder, Christian Debes, Raquel Fandos, Hauke Fath, Tai Fei, Gökhan Gül, Lala Khadidja Hamaidi, Ulrich Hammes, Philipp Heidenreich, Dr. Roy

Howard, Amare Kassaw, Sahar Khawatmi, Renate Koschella, Michael Lang, Stefan Leier, Michael Leigsnering, Zhihua Lu, Marco Moebus, Toufik Mouchini, Ahmed Moustafa, Ivana Perna, Ziliang Qiao, Tim Schäck, Ann-Kathrin Seifert, Waqas Sharif, Sergey Sukhanov, Wassim Suleiman, Gebremichael Teame, Freweyni Teklehaymanot and everyone else whose company I enjoyed during my Ph.D. candidacy.
At this point, I especially thank Michael Fauss and Michael Muma for offering valuable advice and guidance.

Finally, I would like to thank my family and all my friends for being there for me and always believing in me.
My parents, Gisela and Michael, have always supported me in pursuing my personal goals. I sincerely wish to thank them for their love and for their constant, unconditional support in every aspect of my life.

Darmstadt, April 8, 2017

Christian Weiß

# Kurzfassung

In den letzten zwei Jahrzehnten erhielten spärliche Darstellungen vermehrt Aufmerksamkeit in zahlreichen ingenieurswissenschaftlichen Anwednungen. Für eine spärliche (*sparse*) Darstellung eines Signals benötigt man ein Wörterbuch (*dictionary*) aus Basiselementen, welche auffallende und diskriminative Merkmale des Signals beschreiben. Wird das Wörterbuch durch ein analytisches Modell generiert, so hängt seine Aussagekraft von der Qualität dieses Modells ab.
In der vorliegenden Dissertation wird das Problem der Schätzung von spärlichen Darstellungen in der Gegenward von Fehlern und Unsicherheiten bezüglich des Wörterbuchs behandelt. Im ersten Teil wird ein statistisches Rahmenwerk zur spärlichen Regularisierung eingeführt. Der zweite Teil befasst sich mit der Entwicklung von Methodiken zur Schätzung von spärlichen Darstellungen aus hoch-redundanten Wörterbüchern mit unbekannten Wörterbuchparametern. Die vorgestellten Methoden werden mit Hilfe von Anwendungen der Richtungsbestimmung und der faseroptischen Abtastung ausgewertet. Diese dienen als illustrative Beispiele zur Erforschung der abstrakten Probleme in der Theorie der spärlichen Darstellungen.
Das Schtzen einer spärlichen Darstellung schließt oftmals die Lösung eines regularisierten Optimierungsproblems mit ein. Das vorgestellte Rahmenwerk zur Regularisierung bietet ein systematisches Verfahren zur Bestimmung eines Regularisierungsparameters, welcher die Verbundseffekte von Modellfehlern und Messrauschen erfasst. Er wird bestimmt durch eine obere Grenze für den mittleren quadratischen Fehler zwischen den fehlerhaften Daten und dem idealen Modell. Trotz angemessener Regularisierung bleibt die Qualität und die Genauigkeit der erhaltenen spärlichen Darstellung durch die Modellfehler beeinträchtigt und ist durchaus empfindlich bezüglich Änderungen des Regularisierungsparameters. Um dieses Problem zu erleichtern wird eine Wörterbuchkalibrierung durchgeführt. Das Rahmenwerk wird für das Problem der Richtungsschätzung angewendet.
Redundanz ermöglicht es mit dem Wörterbuch eine breitere Klasse von Beobachtungen zu beschreiben. Jedoch wird dadurch auch die Ähnlichkeit zwischen den Einträgen erhöht, was zu missverständlichen Darstellungen führt. Um das Problem der Redundanz und der zusätzlichen Unsicherheit im Wörterbuch zu behandeln werden zwei Strategien verfolgt. Zunächst wird eine Methode der alternierenden Schätzung zur iterativen Bestimmung der zugrundeliegenden spärlichen Darstellung und der Wörterbuchparameter vorgestellt. Ferner werden theoretische Grenzen für den Schätzfehler hergeleitet. Zweitens wird ein Bayes'sches Rahmenwerk zum Schätzen von spärlicher Darstellungen und zum Lernen von Wörterbüchern entwickelt. Ein hierarchisches Modell wird betrachtet um Unsicherheiten in vorherigen Annahmen zu erfassen. Das betra-

chtete Modell für die Koeffizienten der spärlichen Darstellung ist speziell für dem Umgang mit Redundanz im Wörterbuch konzipiert. Approximative Inferenz wird unter Verwendung eines hybriden Markov Chain Monte Carlo Algorithmus durchgeführt. Die Leistungsfähigkeit und die praktische Anwendbarkeit beider Methodiken wird für ein Problem aus der faseroptischen Abtastung ausgewertet. Dabei wird ein analytisches Modell für das Sensorsignal zusammengestellt, welches dann genutzt wird um ein geeignetes Wörterbuch zu generieren.

# Abstract

In the last two decades, sparse representations have gained increasing attention in a variety of engineering applications. A sparse representation of a signal requires a dictionary of basic elements that describe salient and discriminant features of that signal. When the dictionary is created from a mathematical model, its expressiveness depends on the quality of this model.
In this dissertation, the problem of estimating sparse representations in the presence of errors and uncertainty in the dictionary is addressed. In the first part, a statistical framework for sparse regularization is introduced. The second part is concerned with the development of methodologies for estimating sparse representations from highly redundant dictionaries along with unknown dictionary parameters. The presented methods are illustrated using applications in direction finding and fiber-optic sensing. They serve as illustrative examples for investigating the abstract problems in the theory of sparse representations.
Estimating a sparse representation often involves the solution of a regularized optimization problem. The presented regularization framework offers a systematic procedure for the determination of a regularization parameter that accounts for the joint effects of model errors and measurement noise. It is determined as an upper bound of the mean-squared error between the corrupted data and the ideal model. Despite proper regularization, the quality and accuracy of the obtained sparse representation remains affected by model errors and is indeed sensitive to changes in the regularization parameter. To alleviate this problem, dictionary calibration is performed. The framework is applied to the problem of direction finding.
Redundancy enables the dictionary to describe a broader class of observations but also increases the similarity between different entries, which leads to ambiguous representations. To address the problem of redundancy and additional uncertainty in the dictionary parameters, two strategies are pursued. Firstly, an alternating estimation method for iteratively determining the underlying sparse representation and the dictionary parameters is presented. Also, theoretical bounds for the estimation errors are derived. Secondly, a Bayesian framework for estimating sparse representations and dictionary learning is developed. A hierarchical structure is considered to account for uncertainty in prior assumptions. The considered model for the coefficients of the sparse representation is particularly designed to handle high redundancy in the dictionary. Approximate inference is accomplished using a hybrid Markov Chain Monte Carlo algorithm. The performance and practical applicability of both methodologies is evaluated for a problem in fiber-optic sensing, where a mathematical model for the sensor signal is compiled. This model is used to generate a suitable parametric dictionary.

# Contents

# Chapter 1
# Introduction

The choice of a suitable data representation is an integral aspect of signal processing and machine learning. Depending on the problem at hand, different representations can be used to efficiently describe the data in a compact form and to reveal and extract particular discriminative features that might be hidden otherwise [1, 2].
Basic linear transformations have been widely used to obtain suitable representations. In the 1980's, a lot of research was designated to transformations that capture localized or multiscale features [3]. This also led to the development of popular compression algorithms such as the JPEG2000 algorithm for image compression [4]. In the 1990's, the concept of *dictionaries* emerged, where a representation (or approximation) is obtained by choosing a subset of elements from a larger collection, called dictionary [3]. While dictionaries were initially based on existing transformations, the idea of *sparsity* has initiated the development of more general and efficient dictionaries with strong descriptive power [3]. They paved the way for high-accuracy feature estimation by estimating the best underlying sparse representation with respect to the corresponding dictionary [5]. Rather than using a fixed set of dictionary elements, the aim is to incorporate the data itself to identify the best subset of elements, such that the error between the data and its representation (approximation) is minimized – a general problem that belongs to the field of non-linear approximation theory [6]. This idea has further driven the development of efficient sparse estimation algorithms [7–10].
A sparsity-promoting dictionary that reveals particular features can be created in different ways. It can be designed based on a mathematical model for the underlying physical processes generating the data or it can be learned directly from training data [3]. While both approaches have their advantages and limitations, this work is concerned with the former case. Nevertheless, mathematical models are often used as an approximation to describe complicated natural phenomena. Therefore, they can be inaccurate or contain unknown parameters due to imperfect prior knowledge. As a result, the dictionary itself may contain various sources of uncertainty, which leads to inaccurate sparse representations.
The field of estimating sparse representations covers many interdisciplinary topics in mathematics, physics, and computer science. Therefore, it is highly versatile and has accumulated a broad body of literature. Nevertheless, many questions still remain unanswered. This dissertation addresses the problem of estimating sparse representations from dictionaries that are created from mathematical models, where potential sources of uncertainty arise from idealized assumptions or imperfectly known model parameters.

## 1.1 Motivation

Sparse representations have gained increasing importance and their ubiquitous existence is presumed in various applications such as source localization [5], laser ranging [11], sparse coding [12–15] or imaging [16–19]. They also play a central role in the advent of compressed sensing [20–22] – an efficient signal acquisition scheme that has been widely studied with the aim to abate the sampling rate as well as to reduce storage requirements and even hardware costs. The assumption (or the constraint) of sparsity can be instrumental to obtain a unique solution of underdetermined linear equation systems, which are omnipresent in general engineering applications [23]. It has been shown that sparse representations are suitable to achieve super-resolution and to estimate quantities at high accuracy [5]. Nevertheless, this property can only hold under the premise that the model used to create the dictionary correctly describes the observed physical processes that generate the data.
In many practical problems, the existence of unknown model parameters or model errors is inevitable, e.g. due to a limited fabrication accuracy in the manufacturing process of devices or due to external factors such as temperature fluctuations or erosion. Estimating a meaningful sparse representation that correctly highlights the features of interest becomes non-trivial in the presence of dictionary defects or uncertainty. Finding and evaluating solutions for this problem is of general interest.
In order to obtain highly sparse representations, redundancy in the dictionary is imperative [3]. It improves the descriptive power, such that complicated natural phenomena can be described by a single dictionary [9, 23, 24]. However, a high level of redundancy increases the similarity between different dictionary elements, which can cause problems in sparse estimation [25]. Methodologies for sparse estimation that are able to obtain meaningful and accurate sparse representations from highly redundant dictionaries with uncertainty are strongly desirable and of broad practical interest.
The addressed research topics are initially motivated by application-specific problems in direction finding and compressed fiber-optic sensing. Therefore, these applications serve as illustrative examples and are representative for investigating the general and rather abstract problems in the theory of sparse representations. In fact, the problem of direction finding is fundamental and arises explicitly or implicitly in communications, imaging and radar/sonar, just to name a few. For fiber sensing, there exist indeed many sensing technologies. In the scope of this thesis, optical sensors with compressed sensing-based signal acquisition are considered, where the main objective is to estimate the reflection delays of the interrogating signal. This can be generalized to estimating the translation coefficients of localized signals – a frequently encountered problem in different applications, e.g. in laser ranging [11] or radar [19].

## 1.2 Aims and Research Objectives

The aim of this dissertation is the design and development of systematic strategies for estimating sparse representations in the presence of uncertainty or errors in the underlying model that is used to generate the dictionary. The major research objectives are concerned with two particular aspects of this problem as stated below. These are meant to be investigated by means of applications that are representative for frequently encountered problems in research and of broad interest in industry. In particular, direction finding and fiber-optic sensing combined with compressed sensing are in the focus of this work. The research objectives considered in this dissertation are:

1. **Sparse regularization and dictionary calibration:**
   The objective is the development of a statistical framework for sparse regularization in the presence of model errors. Estimating a sparse representation often involves the solution of a regularized optimization problem. A suitable regularization parameter, however, is usually non-trivial to obtain and often hard to justify, especially when an idealized model is used to describe a complicated phenomenon. A common approach is to repeatedly solve the optimization problem for a large range of regularization parameter values at the cost of a high computational load. A statistical regularization framework, in contrast, is intended to yield a suitable regularization parameter prior to sparse estimation, which can be justified by statistical arguments. Since regularization alone can only adjust the tolerance level, the sparse solution can still be inaccurate. Moreover, it is not guaranteed that a solution obtained from an inaccurate model is also a desirable solution. Therefore, in addition to regularization, dictionary calibration is investigated with the aim to improve the quality of the sparse representation and to reduce the sensitivity to the regularization parameter.

2. **Sparse estimation and dictionary learning with highly redundant dictionaries:**
   The objective is to design sparse estimation and dictionary learning methods that are able to accurately estimate an underlying sparse representation along with uncertain dictionary parameters when the entries of the dictionary exhibit high similarity and contain unknown parameters. Redundancy favors highly sparse representations and, in principle, improves the accuracy at which quantities can be estimated. However, the requirements for unique sparse estimation are no longer fulfilled and a desired underlying representation cannot be retrieved using common sparse estimation algorithms. The problem of sparse estimation and

dictionary learning is often addressed by methods that follow the alternating estimation paradigm or by developing a whole Bayesian framework. The goal is to develop estimation frameworks for both methodologies that are designed to estimate an underlying sparse representation from highly redundant dictionaries along with unknown dictionary parameters. In particular, it is intended to investigate the quality and accuracy of the sparse representation and the dictionary parameters when they are estimated based on a

(a) Deterministic sparse model

(b) Probabilistic sparse model.

## 1.3 Contributions

The main contributions of this thesis are listed below.

Contributions related to the first research objective:

- **Statistical sparse regularization framework for direction finding:**
  A statistical sparse regularization framework for direction finding in the presence of model errors and measurement noise is developed. To this end, a direct relationship between the model errors and the regularization parameter of the optimization problem is established. Following a chain of statistical arguments, the regularization parameter is obtained as an upper bound of the mean-squared error between the corrupted data and the idealized model. Different from previous approaches, regularization jointly accounts for measurement noise and particular model errors with pre-specified statistics. Also, it is not necessary to repeatedly solve the optimization problem for a range of parameters, which saves computational costs. It is further shown how dictionary calibration can yield an accurate stable sparse representation with low sensitivity to the regularization parameter. The performance and computational complexity is evaluated in comparison to competing methods in various scenarios.

Contributions related to the second research objective:

- **Sparse estimation/dictionary learning framework for fiber sensing:**
  A versatile sparse estimation and dictionary learning framework for fiber-optic sensing with compressed sensing-based acquisition is presented. The efficient signal acquisition scheme reduces the average sampling rate and the number of samples to be stored and processed. For the sensor signal, a generic model is provided, which is adapted to match real sensor data for experimental validation. Based on this model, a suitable parametric dictionary for high-resolution sensing is created. The 'coherence distance' is introduced as a measure of redundancy for general dictionaries with comparable structure. Different from conventional measures, it emphasizes the similarity of dictionary entries within structured subsets for different parametrizations. Theoretical error bounds are derived for jointly estimating deterministic sparse coefficients and dictionary parameters. They are estimated based on alternating estimation in combination with dictionary pre-processing to handle strong similarity between the entries of the dictionary. The computational complexity of this method is analyzed. The estimation performance is evaluated in simulations, while the practical applicability is verified using real fiber sensing data.

- **Probabilistic model for sparse estimation of localized signals:**
  A probabilistic model for the coefficients of the sparse representation and the dictionary parameters is presented. It is specified for the application of compressed fiber sensing but can be generally adopted for estimating the translation coefficients of localized signals. A hierarchical model structure accounts for uncertainty in prior assumptions. Instead of performing dictionary pre-processing, the problem of redundancy is addressed by leveraging the dictionary structure to promote additional collective shrinkage based on a local similarity model. An alternative interpretation of the sparse model is provided in terms of non-convex optimization. Approximate inference is accomplished using a hybrid Markov Chain Monte Carlo algorithm. Dictionary learning is performed by (i) using a variant of alternating estimation and (ii) sampling the posterior distribution of the parameters. Simulations are used to provide a comparative performance analysis between the probabilistic and the deterministic sparse model, and to highlight advantages and limitations. The practical applicability is verified using real fiber sensing data.

## 1.4 Publications

The following publications have been produced during the period of doctoral candidacy:

**Internationally Refereed Journals Articles**

- C. Weiss and A. M. Zoubir, "Dictionary Learning Strategies for Compressed Fiber Sensing Using a Probabilistic Sparse Model," submitted to *IEEE Transactions on Signal Processing*, 2016.
- C. Weiss and A. M. Zoubir, "A Compressed Sampling and Dictionary Learning Framework for Wavelength-Division-Multiplexing-Based Distributed Fiber Sensing," accepted for publication in *Journal of the Optical Society of America A*, 2017 (assigned issue: vol. 34, no. 5).

**Internationally Refereed Conference Papers**

- C. Weiss and A. M. Zoubir, "Fiber Sensing Using Wavelength-Swept Lasers: A Compressed Sampling Approach," in *Proc. of the 3rd International Workshop on Compressed Sensing Theory and its Applications to Radar, Sonar and Remote Sensing (CoSeRa)*, Pisa, Italy, June 2015. [ *Best Student Paper Award, 1st price* ].
- C. Weiss and A. M. Zoubir, "Robust High-Resolution DOA Estimation with Array Pre-Calibration," in *Proc. of the 22nd European Signal Processing Conference (EUSIPCO)*, Lisbon, Portugal, September 2014.
- C. Weiss and A. M. Zoubir, "DOA Estimation in the Presence of Array Imperfections: A Sparse Regularization Parameter Selection Problem," in *Proc. of the IEEE Workshop on Statistical Signal Processing (SSP)*, Gold Coast, Australia, June/July 2014.
- C. Weiss A. M. Zoubir, "A Sparse Regularization Technique for Source Localization with Non-uniform Sensor Gain," in *Proc. of the IEEE 8th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, A Coruña, Spain, June 2014. [ *Nominated for the Best Student Paper Award* ]
- C. Weiss and A. M. Zoubir, "Fiber Sensing Using UFWT-Lasers and Sparse Acquisition," in *Proc. of the 21st European Signal Processing Conference (EUSIPCO)*, Marrakech, Morocco, September 2013.
- C. Debes, C. Weiss and A. M. Zoubir, "How Good is Your Super-Resolution Image? Quality Assurance in Image Reconstruction Using the Bootstrap," in *Proc. of the 38th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Vancouver, Canada, May 2013.

## 1.5 Overview

The organization of this thesis and a chapter outline are given below.

In **Chapter 2**, the relevant fundamentals of sparse estimation and dictionaries are reviewed. Section 2.1 focuses on sparse representations. Besides common types of dictionaries, the task of dictionary learning is introduced. In Section 2.2, theoretical reconstruction guarantees for sparse estimation with redundant dictionaries are introduced. Also, coherence measures for redundant dictionaries are presented. The contribution of this chapter is the 'coherence distance' for general translation-invariant dictionaries.

In **Chapter 3**, the applications that are relevant in the scope of this thesis and the concept of compressed sensing are introduced. Direction finding is explained in Section 3.1, while fiber-optic sensing is described in Section 3.2. Compressed sensing is introduced in Section 3.3. The contribution of this chapter is the definition of a translation-invariant dictionary for fiber-optic sensing along with a mathematical description of the sensing process in the context of sparse estimation.

In **Chapter 4**, the statistical sparse regularization framework for direction finding is introduced. While Section 4.1 introduces and motivates the topic of sparse regularization, Section 4.2 offers an overview of relevant state-of-the-art methods and concepts. In Section 4.3, the considered model errors are detailed and the regularization problem is introduced in the context of constrained optimization. Section 4.4 and 4.5 introduce the proposed regularization methods for two kinds of model errors. Dictionary calibration is presented in Section 4.6. Section 4.7 analyzes the computational complexity and the results and findings are discussed in Section 4.8. Some concluding remarks are given in Section 4.9. The contribution of this chapter is the above-mentioned statistical sparse regularization framework with dictionary calibration.

In **Chapter 5**, the sparse estimation and dictionary learning framework for fiber sensing is presented. Section 5.1 introduces the task of parametric dictionary learning and the concept of alternating estimation. In Section 5.2, related work and state-of-the-art techniques are reviewed. Section 5.3 describes the considered problem. Theoretical performance bounds for the particular problem at hand are derived in Section 5.4. In Section 5.5, the unified framework for compressed fiber sensing and dictionary learning is introduced. It includes a description of the system architecture and provides a detailed model for the received sensor signal from which the dictionary for sparse estimation is created. In addition, a sparse estimation and dictionary learning algorithm

for highly coherent dictionaries is presented along with an analysis of the computational complexity. In Section 5.7, the performance is evaluated in simulations and experimental validation is provided using real data. A discussion of the results and findings is given in Section 5.8, while Section 5.9 concludes this chapter. The contribution of this chapter is the above-mentioned sparse estimation and dictionary learning framework for fiber sensing using highly coherent dictionaries.

In **Chapter 6**, a probabilistic model for sparse estimation and dictionary learning with localized signals is presented. Section 6.1 introduces the estimation task along with related topics such as prior selection and approximate inference. Section 6.2 provides an overview of state-of-the-art methods and related work. A description of the considered problem is given in Section 6.3. Section 6.4 introduces a sparse model for estimating the translation coefficients of localized signals, including a local covariance model to deal with high redundancy in the dictionary. Also, a relation to non-convex optimization with $\ell_p$-norm constraints is established. Section 6.5 given an overview of the complete hierarchical model and introduces the hybrid Markov Chain Monte Carlo method used for inference in this model. Section 6.6 presents a hybrid and a full Bayesian strategy for dictionary learning. In Section 6.7, the performance is evaluated using simulations and real data. A comparison between the results obtained using the presented probabilistic sparse model and the deterministic sparse model in Chapter 5 is provided. The results and findings are discussed in Section 6.8 and Section 6.9 gives a conclusion for this chapter. The contribution of this chapter is the above-mentioned probabilistic model for sparse estimation of localized signals with highly coherent dictionaries.

**Chapter 7** provides a summary and a conclusion for this dissertation along with an outlook of some future research directions.

# Chapter 2

# Primer on Sparse Estimation

Sparse signals have gained increasing attention in various signal processing and machine learning applications. These include, for example, imaging applications such as synthetic aperture radar imaging [18, 19] or medical imaging [16, 17], sparse coding [12–15], source localization [5], direction-of-arrival estimation and laser ranging [11]. Compressed sensing [20, 21, 26] is another popular application in sparse signal processing, where the sparse structure of signals is exploited to alleviate the constraints and requirements imposed on an acquisition system in terms of storage, sampling rate, and hardware costs [20, 22, 26, 27].

The assumption of sparsity can help to solve underdetermined linear equation systems, which are of central importance and often found in engineering applications [23]. It is well-known that most signals encountered in the wild are not exactly sparse but contain only a few significant components. Such signals are called 'compressible' [20]. However, they are usually not directly sparse or compressible in the sensing domain. The sparse structure is rather revealed in terms of an appropriate 'dictionary', which can represent the signal using only a few non-zero coefficients. Hence, the dictionary is a key component in sparse signal processing and composed of atomic elements that efficiently capture the structure of a certain class of signals. Following the notion in [8, 28], the entries of a dictionary are referred to as 'atoms'. When a larger class of signals is covered by a single dictionary, the resulting representations are surely less sparse than in cases, where the dictionary is specifically designed for a certain sub-class of signals (cf. redundant dictionaries in [29] and below).

The primary objective in sparse signal processing is the construction or identification of suitable dictionaries as well as the development of intelligent algorithms, that can efficiently exploit the sparse structure of the signal in order to extract the desired information. There exist various 'off-the-shelf'-dictionaries that are applicable to a wide class of signals, e.g. Fourier or wavelet dictionaries [26]. Identifying a dictionary for a particular problem is not a trivial task and requires sufficient knowledge about the signal structure and the physical processes generating the signals.

In this chapter, the foundations of sparse signals, different classes of dictionaries, sparse inverse problems, sparse recovery algorithms as well as some popular reconstruction guarantees and quality measures are briefly reviewed. They serve as prerequisites for the ensuing chapters.

The definition of the coherence distance in Section 2.2.1.4 is partly taken from [30][1].

**Chapter Outline**

Section 2.1 introduces the concept of sparse representations and describes the properties of common types of dictionaries. Also, dictionary learning is introduced. In Section 2.2, the problem of sparse estimation from redundant dictionaries is discussed, along with theoretical reconstruction guarantees and coherence measures.

## 2.1 Sparse Representation of Signals

The fundamental assumption in sparse signal processing is, that an observed or measured signal can be sparsely represented with respect to a certain dictionary. Subsequently, a signal is described in terms of a generalized vector over the field of real or complex numbers in a Hilbert space, endowed with an inner product. In particular, any observed analog signal is modeled as a member of the Hilbert space $\mathcal{L}^2$ of square-integrable functions. The focus of this thesis is the analysis of discrete signals, that can be represented by finite-dimensional vectors, $\mathbf{r} \in \mathbb{C}^L$, in the Euclidean space. The signal $\mathbf{r}$ may represent the Nyquist samples of an observed analog signal. In this context, a 'dictionary', $\mathbf{D} = [\mathbf{d}_1, \dots, \mathbf{d}_N]$, is a collection of vectors (atoms), $\mathbf{d}_i \in \mathbb{C}^L, i = 1, \dots, N$, that form a basis or represent a redundant set of vectors, such that $\operatorname{span}(\mathbf{d}_1, \dots, \mathbf{d}_N) = \mathbb{C}^L$, $N \geq L$.
In order to obtain a sparse representation, $\mathbf{r}$ is decomposed into a weighted sum of dictionary atoms, that is

$$\mathbf{r} = \sum_{i=1}^{N} x_i \, \mathbf{d}_i \;=\; \mathbf{D}\mathbf{x}\,, \tag{2.1}$$

where most of the coefficients in $\mathbf{x} = [x_1, \dots, x_N]^\top$ are zero (or close to zero for compressible signals). The set $\mathcal{S}$ is defined to contain the indices of the $K$ non-zero elements in $\mathbf{x}$, i.e. the sparse support $\mathcal{S} = \operatorname{supp}(\mathbf{x})$ (or the indices of the $K$ most significant components in $\mathbf{x}$ for compressible signals). Throughout this chapter, real-valued signals and dictionaries are considered. In the complex-valued case, the signals and the dictionary can be split up into real and imaginary parts.

[1] C. Weiss and A. M. Zoubir, "A Compressed Sampling and Dictionary Learning Framework for Wavelength-Division-Multiplexing-Based Distributed Fiber Sensing," accepted for publication in *Journal of the Optical Society of America A*, 2017 (assigned issue: vol. 34, no. 5).

### 2.1.1 Classes and Properties of Dictionaries

There exist different classes of dictionaries with individual properties. The choice of a certain dictionary usually depends on the application. Some important dictionaries and selection criteria are listed below.

#### 2.1.1.1 Orthogonal Dictionaries

Orthogonal dictionaries are convenient in that the coefficients of a sparse representation can be obtained by a simple projection with the dictionary matrix, i.e. $\mathbf{x} = \mathbf{D}^\top \mathbf{r}$. If the dictionary atoms are also normalized, the dictionary itself describes a unitary transformation. Despite the convenient analytical properties of orthogonal dictionaries, the single atoms suffer from a limited descriptive power [3]. When specific characteristics of the signal are taken into account during the dictionary design process, more efficient sparse representation can be obtained.

#### 2.1.1.2 Redundant Dictionaries

When the number of dictionary atoms is larger than the dimensionality of the observed signal, i.e. $N > L$, the dictionary is called 'redundant'. Redundant dictionaries have two major advantages. Firstly, they can yield much sparser representations than orthogonal dictionaries [3, 9, 12, 29, 31]. Secondly, redundancy allows for high-resolution estimation of quantities directly in the sparse domain [5, 32]. However, a drawback of redundant dictionaries is, that there may exist several ambiguous representations of a signal with the same number of non-zero coefficients. Hence, due to the coherence of the dictionary atoms, the requirements for stable and robust sparse reconstruction may not be fulfilled [9, 25, 27].
In cases, where a signal emerges from different underlying physical processes, its complexity can no longer be captured by a single transformation. Such signals, e.g. sinusoids and spikes, are more appropriately described in terms of mixtures of transformations or dictionaries. Such compositions of different sub-dictionaries are called 'composite dictionaries' [9, 24, 27, 33].

#### 2.1.1.3 Parametric Dictionaries

Parametric dictionaries are created using basic mathematical functions that are often motivated by a mathematical model to describe the underlying physical processes

generating the signal. Therefore, the dictionary atoms are sometimes referred to as 'parametrized waveforms' [8, 34]. Their analytical form allows for an efficient implementation and offers the possibility to obtain optimality proofs and error bounds [3]. Further advantages are storage-efficiency and scalability, since only the parameters need to be stored rather than all samples of each atom. Hence, large-scale dictionaries and arbitrarily-sized signals can be effectively represented [3, 35]. Popular parametric dictionaries are based on linear transforms, e.g. the Fourier or Wavelet transform [26]. The physical motivation of a Fourier dictionary is to describe a signal in terms of its frequency components, while a Wavelet dictionary can be used to describe the features of an image over multiple scales. However, the performance of parametric dictionaries is limited by the assumptions made in order to obtain a mathematical model for the signal. This limitation becomes apparent when many simplifications are necessary to describe a natural phenomenon in terms of a convenient mathematical model [3]. For example, a Fourier dictionary is well suited for representing signals that are generated by a finite sum of monochromatic signals. This is no longer true if the observed signal also contains transient components. If an accurate generative signal model is available, parametric dictionaries can efficiently match the structure of certain signals [35].
One important class of parametric dictionaries are 'translation-invariant dictionaries'. They are created from a parametric function and the dictionary atoms correspond to distinct translations of that function. Such dictionaries are useful in estimating the translation coefficients of localized signals [36–39].

#### 2.1.1.4 Non-parametric Dictionaries

Some complicated natural phenomena cannot be described in terms of a convenient mathematical model. This problem has motivated the idea of constructing the dictionary directly from the data based on a training set. It has become a prominent technique in the advent of Machine Learning. Dictionaries constructed in this way can adapt better to specific realizations of the data [3, 40].

#### 2.1.1.5 Dictionaries With Uncertainty

In many practical problems, there exist several sources of uncertainty. They may arise from imperfections in the manufacturing process of devices, errors and incorrect assumptions in the underlying model or incomplete knowledge of the observed signal. Uncertainty can be incorporated in the dictionary in order to account for these effects. When a parametric dictionary is employed, the basic signal structure is usually known

and uncertainty is reflected in terms of unknown or imperfectly known parameters. One major objective in the ensuing chapters is to estimate such parameters along with a sparse representation of the signal.

### 2.1.2 Dictionary Learning

When a dictionary contains uncertainty, such as unknown parameters, or if the dictionary is *a priori* unknown, the missing information can be extracted from the data under appropriate assumptions and constraints. This particular problem is addressed in *Dictionary Learning* (DL), although the term 'learning' sometimes refers to creating non-parametric dictionaries purely from the data, rather than using the data to estimate parameters of an analytical (parametric) dictionary [3].
Non-parametric DL methods do not make use of an analytic form. They cannot exploit available information about the dictionary structure and rely on general constraints such as the column norms and inter-atom correlation [35,41]. Moreover, purely data-based dictionaries suffer from unknown inner structure and lack of interpretability. Some popular non-parametric DL methods include the Method of Optimal Directions [42,43], the generalized principal component analysis [44], the family of iterative least-squares DL algorithms [45], the K-SVD algorithm [46], and the Recursive Least-Squares method [47]. In addition, there also exist robust approaches [48,49].
Parametric DL methods adapt the parameters of an analytic (parametric) dictionary according to the acquired data. They benefit from the advantages of parametric dictionaries, such as scalability, efficient implementation and low storage costs, but the performance also depends on the quality of the underlying physical model [3,35]. The class of translation-invariant dictionaries is often considered in parametric DL [3,36,37,39].
In order to estimate the parameters of a dictionary given the acquired data, statistical methods, such as *maximum likelihood* (ML) estimation [3, 50] or *maximum a posteriori* (MAP) estimation [3, 51], can be applied.
In order to benefit from the advantages of both approaches, hybrid DL techniques involve the data-driven training of structured parametric dictionaries. This has several advantages in reducing the amount of training data, improving the convergence speed, reducing the density of local minima and offering a more compact representation [3,37,38,45].
*Alternating estimation* (AE) is a prominent heuristic for iteratively solving the resulting optimization problem in order to estimate both, the dictionary (parameters) and the sparse representation of the signal [52,53]. AE is also known as 'Block-nonlinear Gauss-Seidel method' or 'Block coordinate descent method' [54]. The *Expectation Maximization* (EM) algorithm [3,55] is one variant of AE-based estimation, which is often

used in a hybrid Bayesian setting to iteratively estimate unknown model parameters. It represents an alternative to a full Bayesian formulation, in which the dictionary parameters are jointly estimated together with all other latent variables in the model [56]. In the ensuing chapters, AE-based and full Bayesian strategies for parametric DL are pursued.

## 2.2 Sparse Estimation & Redundant Dictionaries

In the theory of sparse representations and redundant dictionaries, one distinguishes two different concepts: 'sparse analysis' and 'sparse synthesis' [3,12,57]. Sparse analysis attempts to represent the data in terms of a linear transformation, say $\mathbf{T}$, such that the outcome $\mathbf{Tr}$ is sparse. Therefore, $\mathbf{T}$ is also called the 'synthesis operator' [12]. When the dictionary, $\mathbf{D}$, forms a 'tight frame', such that $\mathbf{r} = \mathbf{D}\mathbf{D}^\top\mathbf{r}$ (see e.g. [58]), then $\mathbf{D}$ and $\mathbf{T}$ are related by $\mathbf{D}^\top = \mathbf{T}$ and the data is represented by inner products with the dictionary atoms [3]. The sparse synthesis approach, in turn, aims at representing $\mathbf{r}$ in terms of a linear combination with the dictionary atoms, i.e. $\mathbf{r} = \mathbf{D}\mathbf{x}$ as in (2.1). Both representations are equivalent for orthogonal dictionaries but not for redundant dictionaries [3].
In this thesis, the synthesis approach is addressed. It is further assumed, that the signal of interest, $\mathbf{r}$, can not be observed directly. Only a noisy version, $\mathbf{y}$, is available, which leads to the following general signal model:

$$\mathbf{y} \;=\; \mathbf{r} + \mathbf{n} \;=\; \mathbf{D}\mathbf{x} + \mathbf{n}\;, \tag{2.2}$$

where $\mathbf{n}$ is an additive noise component and $\mathbf{D}$ represents a redundant dictionary. The goal is to estimate a sparse representation, $\mathbf{x}$, that minimizes a certain cost function, $C(\mathbf{x})$. Without the sparsity assumption, there exist infinitely many possible representations $\mathbf{x} + \Delta\mathbf{x}$, where $\Delta\mathbf{x}$ lies in the null-space of $\mathbf{D}$. Therefore, sparsity also helps to obtain a unique representation. The problem of finding the sparsest solution in (2.2) belongs to the field of 'highly non-linear approximation theory' [6].
A common choice for the cost function is the squared error, i.e. $C(\mathbf{x}) = \|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2^2$, and sparsity can be enforced by minimizing the $\ell_0$-norm [2]. This leads to a combinatorical problem, which is intractable for high-dimensional vectors, $\mathbf{x}$. Under certain conditions (discussed below), an equivalent sparse solution can be obtained using the $\ell_1$-norm instead [28,59]. Hence, since the $\ell_1$-norm represents the tightest convex relaxation of the

[2] Strictly speaking, the term 'norm' is reserved for operations that fulfill the original axioms of a true norm. Therefore, operations such as $\|\mathbf{x}\|_0$ should rather be called a 'pseudo-norm'. However, due to the relaxed usage of this term in the general literature of sparse estimation, it is adopted throughout this thesis to describe the general class of operations $\|\cdot\|_p$, with $0 \leq p \leq 2$.

$\ell_0$-norm, the powerful machinery of convex optimization becomes available for solving the problem. Written in a parametric form, the resulting optimization problem resembles the LASSO problem in [60], i.e.

$$\arg\min_{\mathbf{x}} \ \|\mathbf{y}-\mathbf{D}\mathbf{x}\|_2^2 \ + \ \lambda_L\|\mathbf{x}\|_1\,, \tag{2.3}$$

where $\lambda_L$ is a regularization parameter. The solution of this problem also corresponds to an MAP estimation problem (c.f. e.g. [61]) with a Gaussian likelihood,

$$p(\mathbf{y}\,|\,\mathbf{x}) \ \propto \ \exp\!\left(-\frac{1}{2\sigma_n^2}\|\mathbf{y}-\mathbf{D}\mathbf{x}\|_2^2\right), \tag{2.4}$$

and i.i.d. zero-mean Laplace-distributed elements $x_i$, i.e.

$$p(x_i) \ \propto \ \exp(-\lambda_L|x_i|)\,, \quad i=1,\ldots,N. \tag{2.5}$$

By maximizing the log-posterior, $\log(p(\mathbf{y}\,|\,\mathbf{x})p(\mathbf{x}))$, one obtains (2.3).
Another formulation of the sparse estimation problem is that of '$\ell_1$-minimization' in [9, 25], which is also called 'Basis Pursuit Denoising' [8, 25]. It is given by

$$\min_{\boldsymbol{x}} \|\mathbf{x}\|_1 \quad \text{s.t.} \quad \|\mathbf{y}-\mathbf{D}\mathbf{x}\|_2 \leq \beta_{\text{reg}}\ , \tag{2.6}$$

where $\beta_{\text{reg}}$ is a hyperparameter to account for noise or model errors. Although $\ell_1$-minimization can obtain globally optimal solutions with robustness to measurement errors and noise, the computational complexity of optimization-based methods is often high [25]. Greedy methods, such as OMP [7] or CoSaMP [10], offer another possibility to obtain a sparse solution in (2.2). They usually require lower computational costs but are not guaranteed to obtain the globally optimal solution [25].

### 2.2.1 Reconstruction Guarantees and Coherence Measures

Whether or not the sparse coefficients can be successfully estimated depends on certain joint conditions on the sparse signal, $\mathbf{x}$, and the dictionary, $\mathbf{D}$. Especially when $\mathbf{D}$ is a redundant dictionary, there might exist several equivalent $K$-sparse representations of the original signal, $\mathbf{r}$, such that robust and stable reconstruction might no longer be possible [9, 27, 31]. A popular reconstruction guarantee is provided by the *restricted isometry property* (RIP) [13, 25, 62]. It is defined by

$$(1-\delta_K)\|\mathbf{x}\|_2^2 \leq \|\mathbf{D}\mathbf{x}\|_2^2 \leq (1+\delta_K)\|\mathbf{x}\|_2^2\ , \tag{2.7}$$

where $\delta_K \geq 0$ is called the 'restricted isometry constant'. For a given dictionary, $\mathbf{D}$, and a certain sparsity level of the signal, $K$, the constant $\delta_K$ is computed in order to assess

the conditions for sparse reconstruction. Intuitively, the RIP states that, if $\delta_K$ is sufficiently small, every subset of $s \leq K$ columns approximately resembles an orthonormal system [27]. In the presence of model errors, or if $\mathbf{x}$ is not ideally $K$-sparse, certain stability and robustness conditions can be derived [25]. For many popular sparse estimation algorithms, including $\ell_1$-minimization and OMP, there exist upper bounds, $\delta^*$, that guarantee the success of sparse reconstruction. For OMP, the RIP of order (sparsity level) $K+1$ must be fulfilled with isometry constant $\delta_{K+1} < 1/(3K^{1/2}) = \delta^*$ [63]. For $\ell_1$-minimization, error bounds can be derived for the RIP of order $2K$ with isometry constant $\delta_{2K} < 4/\sqrt{41} = \delta^*$ [25].

It can be shown that 'sub-Gaussian random matrices' fulfill the RIP at a given sparsity level with high probability [20, 25]. A sub-Gaussian random variable, $X_{\mathrm{sg}}$, is defined in terms of some constants, $\beta_{\mathrm{sg}}, \kappa_{\mathrm{sg}} > 0$, such that the probability $\Pr(|X_{\mathrm{sg}}| \geq t) \leq \beta_{\mathrm{sg}} \mathrm{e}^{-\kappa_{\mathrm{sg}} t}$, $\forall\ t > 0$. Likewise, the entries of a sub-Gaussian random matrix are i.i.d. zero-mean sub-Gaussian random variables with variance 1 and sub-Gaussian parameters, $\beta_{\mathrm{sg}}$ and $\kappa_{\mathrm{sg}}$ [25]. Examples are Gaussian matrices themselves, Bernoulli- and Rademacher-matrices [25]. The entries of a Bernoulli matrix take the values $\{0, 1\}$ and those of a Rademacher matrix take values $\{1, -1\}$, both with equal probabilities. A useful reconstruction guarantee for sub-Gaussian matrices is given below:

Let $\mathbf{S}_G \in \mathbb{R}^{M \times L}$ be a sub-Gaussian matrix with sub-Gaussian parameters $\beta_{\mathrm{sg}}, \kappa_{\mathrm{sg}}$, and let $\delta^*$ be a given RIP constant. If $M$ fulfills the relation

$$M \geq C \left(\frac{1}{\delta^*}\right)^2 \left( K \log(e\, N/K) + \log(2\epsilon_{\mathrm{SG}}^{-1}) \right) \tag{2.8}$$

for some constant $C_{\mathrm{sg}} > 0$, that depends only on $\beta_{\mathrm{sg}}, \kappa_{\mathrm{sg}}$, then $\mathbf{S}_G/\sqrt{M}$ satisfies $\delta_K \leq \delta^*$ with probability at least $1 - \epsilon_{\mathrm{SG}},\ \epsilon_{\mathrm{SG}} \ll 1$ [25].

For general dictionaries, however, verifying the RIP is NP-hard and there exist no polynomial-time algorithms, to date, for explicitly constructing a dictionary that meets the RIP requirements [25]. In order to overcome this problem, other computationally tractable quality measures have been proposed. They often rely on the 'coherence' between the dictionary atoms, i.e. the level of similarity between dictionary columns. In the case of redundant dictionaries, matrix pre-conditioning can be used to reduce the dictionary coherence problem [27]. Some popular coherence measures are reviewed below.

#### 2.2.1.1 The Mutual Coherence

The mutual coherence is a popular and widely used criterion to measure and quantify the dictionary coherence between all pairs of atoms [12, 25, 28]. It is defined by

$$\mu_{\text{MC}}(\mathbf{D}) = \max_{j \neq i} |\mathbf{d}_j^{\text{H}} \mathbf{d}_i| = \max_{j \neq i} |[\mathbf{D}^{\top}\mathbf{D}]_{ji}| \ , \quad j, i \in \Omega \ , \tag{2.9}$$

where $\Omega = \{\, i \in \mathbb{N} \,|\, 1 \leq N \,\}$ contains all column indices of the dictionary, $\mathbf{D}$.

#### 2.2.1.2 The Babel Function / $\ell_1$-coherence

The Babel function, also known as $\ell_1$-coherence [9, 12], takes into account the specific subset of $s$ dictionary atoms with indices in $\Lambda_B \subset \Omega$, where $\Lambda_B$ has cardinality $|\Lambda_B| = s$. This subset shows the maximum cumulative correlation with all remaining atoms $\mathbf{d}_i,\ i \in \Omega \setminus \Lambda_B$. The Babel function is defined by

$$\mu_B(\mathbf{D}, s) \ = \max_{|\Lambda_B| = s} \ \max_{i \in \Omega \setminus \Lambda_B} \sum_{j \in \Lambda_B} |\mathbf{d}_i^{\text{H}} \mathbf{d}_j| \ \leq \ s\, \mu_{\text{MC}}(\mathbf{D}) \ . \tag{2.10}$$

It is a non-decreasing function in $s$ and related to the mutual coherence by $\mu_B(\mathbf{D}, 1) = \mu_{\text{MC}}(\mathbf{D})$. If $\mu_B(\mathbf{D}, s)$ increases 'slowly' in $s$, then $\mathbf{D}$ is called 'quasi-incoherent' [12].

#### 2.2.1.3 Spark of a Matrix

According to the definition in [9, 12], the 'spark' of a matrix, spark($\mathbf{D}$), is the smallest number of columns that form a linear dependent set. This is in contrast to the 'rank' of a matrix, which is the largest number of linear independent columns [9, 12]. The spark can be used to ensure the uniqueness of a sparse representation. A representation over $K$ atoms of $\mathbf{D}$ is unique if and only if spark($\mathbf{D}$) $< 1/2$ [9, 12]. Lower bounds for the spark can be obtained based on the Babel function, $\mu_B(\mathbf{D}, s)$ [12].

#### 2.2.1.4 The coherence distance for shift-invariant dictionaries

The presented material for describing the 'coherence distance' is partly taken from [30][3].

[3] C. Weiss and A. M. Zoubir, "A Compressed Sampling and Dictionary Learning Framework for Wavelength-Division-Multiplexing-Based Distributed Fiber Sensing," accepted for publication in *Journal of the Optical Society of America A*, 2017 (assigned issue: vol. 34, no. 5).

Shift-invariant dictionaries are created from a generating mathematical function, e.g. a known signal, and the dictionary atoms represent distinct translations of that function. Then, the dictionary is invariant to translations. Usually, it is also highly redundant. These properties also enable the generation of large-scale dictionaries while keeping the number of parameters small [36–39].
The coherence distance represents a measure to assess the level of difficulty in estimating a unique sparse representation for general translation-invariant dictionaries with strong inter-atom coherence. Common coherence measures are often insufficient for this purpose. For example, similar to the mutual coherence, the Babel function yields the coherence level between the $K$ most similar dictionary atoms. Nevertheless, when $K$ is small, the obtained value can be similar for two highly coherent dictionaries, although their overall coherence level is very different. In addition, the structure of the dictionary plays an important role. Translation invariant dictionaries are used to estimate the translation coefficients of localized signals. In some applications, it is not likely or even impossible that two localized signals appear very close to each another. However, for such dictionaries, the $K$ most similar atoms usually appear in a burst very close to each other and are related by an incremental change in the translation coefficient. The mutual coherence and the Babel function do not consider such structures. Therefore, they are not suitable to describe the difference between two highly coherent translation invariant dictionaries, especially when $K$ is small. The coherence distance, in turn, is designed to emphasize small changes in the coherence level of translation-invariant dictionaries. It is an appropriate measure for general dictionaries of the form (5.1), where the similarity between atoms decreases with increasing index difference. Yet, it is possible to establish a relationship between the two measures as shown below.

A redundant, translation-invariant dictionary, $\mathbf{D}$, is considered. The set of atom indices is given by $\Omega = \{1, \ldots, N\}$. For some $\beta_d \in \mathbb{R}$ with $0 < \beta_d < 1$, the coherence distance is defined by

$$d_c(\mathbf{D}, \beta_d) \;=\; \max_{i \in \Omega} \left\{ \underset{\Delta_{ij} = |i-j|}{\arg\min} \;\; |\mathbf{d}_i^\top \mathbf{d}_j| \;\leq\; \beta_d \, |\mathbf{d}_i^\top \mathbf{d}_i| \;,\; j \in \Omega \right\} \;, \tag{2.11}$$

where $|\mathbf{d}_i^\top \mathbf{d}_i| = 1 \;\forall\; i \in \Omega$ since the dictionary atoms are usually normalized. The factor $\beta_d$ in (2.11), is a threshold for the similarity between subsequent atoms with respect to the atom self-coherence. $d_c(\mathbf{D}, \beta_d)$ is the minimally required index distance for which the atom similarity is reduced by $\beta_d$. Intuitively, it determines the size of the largest burst of coherent atoms in a translation-invariant dictionary, where the atom coherence decreases with increasing index distance.
In order to establish a relationship between the coherence distance and the Babel function, $\mu_B(\mathbf{D}, s+1)$ is calculated for an arbitrary sparsity level $s \in \mathbf{N}_+$. Let $\Lambda_s$ be a set of

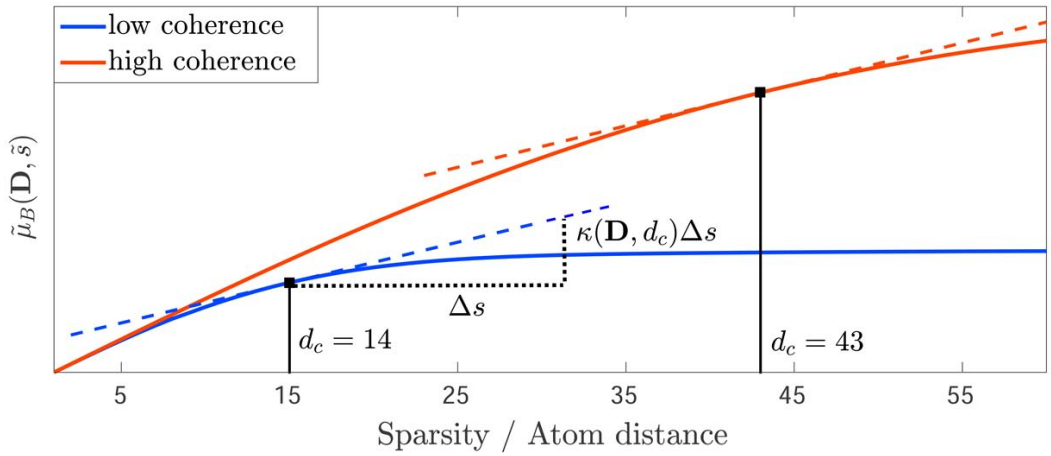



Figure 2.1. Relationship between the Babel function, $\mu_B(\mathbf{D}, s)$, and the coherence distance, $d_c(\mathbf{D})$ with $\beta_d = 0.5$, for high and low coherence levels.

$s+1$ adjacent indices, i.e. $\Lambda_s = \{q \in \Omega \,|\, q_0 \leq q \leq q_0 + s\}$ with $q_0 \in \{2, \ldots, N-s-1\}$, and let $\Lambda_B = \Lambda_s$ in (2.10). Then, the inner maximum in (2.10) is attained by choosing $i^* \in \Omega \backslash \Lambda_s$ directly adjacent to the indices in $\Lambda$, such that $|i^* - q_0| = 1$ or $|q_0 + s - i^*| = 1$. Subsequently, it is assumed that $i^* = q_0 - 1$. Next, $\tilde{\mu}_B(\mathbf{D}, \tilde{s}), \tilde{s} \in \mathbb{R}_+$ is defined to be a continuous extension of $\mu_B(\mathbf{D}, s)$, Using a Taylor series, a first-order approximation can be obtained by

$$\mu_B(\mathbf{D}, s+1) \;=\; \sum_{j=1}^{s+1} |\mathbf{d}_{i^*}^\top \mathbf{d}_{i^*+j}| \;\;\approx\;\; \mu_B(\mathbf{D}, s) + \kappa(\mathbf{D}, s)\Delta s \;, \tag{2.12}$$

where

$$\kappa(\mathbf{D}, s) \;=\; \frac{d\tilde{\mu}_B}{d\tilde{s}}(\mathbf{A}, \tilde{s})|_{\tilde{s}=s} \;. \tag{2.13}$$

Since $\mu_B(\mathbf{D}, s)$ is a concave function, the relation $\kappa(\mathbf{D}, s)\Delta s \geq |\mathbf{d}_{i^*}^\top \mathbf{d}_{i^*+s+1}|$ holds. Then, by setting $\Delta s = 1$, the relationship between $d_c(\mathbf{D}, \beta_d)$ and $\mu_B(\mathbf{D}, s+1)$ is established by

$$d_c(\mathbf{D}, \beta_d) = \;\; \underset{d \in \Omega}{\arg\min} \left\{ \kappa(\mathbf{D}, d) \;\leq\; \beta_d \, |\mathbf{d}_{i^*}^\top \mathbf{d}_{i^*}| \,, \;\; i^* + d_c \leq N \right\}. \tag{2.14}$$

Thus, $d_c(\mathbf{D}, \beta_d)$ can be interpreted as the sparsity level, $s$, for which the Babel function, $\tilde{\mu}_B(\mathbf{D}, \tilde{s})$, grows at a rate $\kappa(\mathbf{D}, d_c(\mathbf{D}, \beta_d)) \leq \beta_d \, |\mathbf{d}_{i^*}^\top \mathbf{d}_{i^*}| < \kappa(\mathbf{D}, d_c - 1)$. The relationship between the slope of $\tilde{\mu}_B(\mathbf{A}, \tilde{s})$ and $d_c(\mathbf{D}, \beta_d)$ for $\beta_d = 0.5$ is depicted in Figure 2.1 for a dictionary with high and low coherence, respectively. In the scope of this thesis, it is assumed that $\beta_d = 0.5$, such that $d_c$ corresponds to the index difference, where the atom coherence is reduced by 3 dB with respect to the self-coherence, $|\mathbf{d}_{i^*}^\top \mathbf{d}_{i^*}|$, for any $i^* \in \{1, \ldots, N-s-2\}$. Therefore, the notation $d_c = d_c(\mathbf{D}, \beta_d)$ is subsequently used when dictionary is clearly specified. At low sparsity levels such as $\tilde{s} < 10$, the slope

of $\tilde{\mu}_B(\mathbf{A}, \tilde{s})$ in Figure 2.1 can merely be distinguished. Thus, it yields no information of the burst size of coherent atoms. However, $d_c$ takes very different values when $s$ is small and emphasizes the difference.

# Chapter 3
# Applications and Signal Models

In this chapter, the relevant applications and corresponding basic signal models are introduced. The applications are representative for frequently encountered estimation problems. They are used to instanciate the abstract formulation of the considered problems in sparse estimation and to evaluate different sparse estimation techniques. In particular, direction finding, fiber-optic sensing and the concept of compressed sensing are introduced.

**Chapter Outline**

Section 3.1 describes the application of direction finding and the considered signal model. In Section 3.2, fiber-optic sensing is introduced, including the sensing principle and a basic signal model. Section 3.3 details the concept of compressed sampling, which can be combined with either application.

## 3.1 Direction Finding

Direction finding aims at localizing the source of a signal and has a wide range of applications, e.g. in radar, sonar and communications [64, 65]. Direction-of-arrival estimation is used to estimate the bearing of a source, while this information can be further utilized to determine its actual location [64].

### 3.1.1 Signal Model

Parts of the presented signal model appear in [66][1], [67][2], [68][3]. and are based on the theoretical foundations in [64, 65].

[1] C. Weiss A. M. Zoubir, "A Sparse Regularization Technique for Source Localization with Non-uniform Sensor Gain," in *Proc. of the IEEE 8th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, A Coruña, Spain, June, 2014.

[2] C. Weiss and A. M. Zoubir, "DOA Estimation in the Presence of Array Imperfections: A Sparse Regularization Parameter Selection Problem," in *Proc. of the IEEE Workshop on Statistical Signal Processing (SSP)*, Gold Coast, Australia, June/July, 2014.

[3] C. Weiss and A. M. Zoubir, "Robust High-Resolution DOA Estimation with Array Pre-Calibration," in *Proc. of the 22nd European Signal Processing Conference (EUSIPCO)*, Lisbon, Portugal, September, 2014.

A uniform linear array of $L$ sensors with sensor spacing $\Delta d$ is considered to estimate the incident direction of monochromatic signals. In order to avoid spatial aliasing and to meet the spacial sampling requirements, the sensor spacing is set to half of the signal wavelength, i.e. $\Delta d = \lambda_0/2$, where $f_0 = c_p/\lambda_0$ is the frequency of the signal and $c_p$ is the speed of propagation. It is further assumed, that the array is aligned with the $x$-axis of a Cartesian coordinate system, according to Figure 3.1. The total signal is modeled as a superposition of plane waves with zero-mean, circular-symmetric complex Gaussian distributed amplitudes of variance $\sigma_u^2$, i.e. $u_k \sim \mathcal{CN}(0, \sigma_u^2)$, $k = 1, \ldots, K$. At the $l$-th sensor, the signal arrives with a delay that depends on the sensor position: $\tau_l = (l-1)\Delta d \cos(\vartheta)/c_p$, $l = 1, \ldots, L$. The variable $\vartheta \in [0, \pi]$ denotes the incident angle of the signal measured in terms of the azimuth angle of the array (c.f. Figure 3.1). The phase-shifts induced by the different delays at the individual sensors are expressed relative to the phase measured at the first sensor with index $l = 1$, i.e.

$$\overline{\phi}_l(\vartheta) \;=\; \omega_0 \tau_l \;=\; 2\pi f_0 \tau_l \;=\; \frac{2\pi}{\lambda_0}(l-1)\Delta d \cos(\vartheta), \quad l = 1, \ldots, L\,. \tag{3.1}$$

For a monochromatic source with a certain bearing, the array response is described in terms of the 'array steering vectors', $\mathbf{a}(\vartheta) = [a_1(\vartheta), \ldots, a_L(\vartheta)]^\top$. They are also called 'array manifold vector' and depend on the incident direction of the signal, $\vartheta$. Their entries are given by

$$a_l(\vartheta) \;=\; \exp\bigl(-\overline{\phi}_l(\vartheta)\bigr) \;=\; \exp\!\left(-\frac{2\pi}{\lambda_0}\,(l-1)\Delta d \cos(\vartheta)\right), \quad l = 1, \ldots, L\,. \tag{3.2}$$

A 'snapshot' of the measurements from all sensors can be written by

$$\mathbf{y} = \sum_{k=1}^{K} u_k\, \mathbf{a}(\vartheta_k) + \mathbf{n}\,, \tag{3.3}$$

where $\mathbf{n}$ is a zero-mean, circular-symmetric complex Gaussian distributed Gaussian noise vector with i.i.d. entries $n_l \sim \mathcal{CN}(0, \sigma_n^2)$, $l = 1, \ldots, L$.
In order obtain a sparse representation of this signal, the model in [5] is adopted, where the dictionary represents a collection of closely spaced steering vectors, i.e.

$$\mathbf{D} = [\mathbf{a}(\tilde{\vartheta}_1), \ldots, \mathbf{a}(\tilde{\vartheta}_N)]\,. \tag{3.4}$$

Assuming that the true DOAs coincide exactly with the angular grid specified in $\mathbf{D}$, one may equivalently write (3.3) in terms of a sparse signal, $\mathbf{x}$, i.e

$$\mathbf{y} = \mathbf{D}\mathbf{x} + \mathbf{n}\,. \tag{3.5}$$

Herein, the non-zero elements, $x_{i_k} = u_k$, appear at the positions $i_k \in \mathcal{S} = \text{supp}\{\mathbf{x}\}$, that correspond to the true DOAs, $\vartheta_k$, $k = 1, \ldots, K$, according to (3.2).

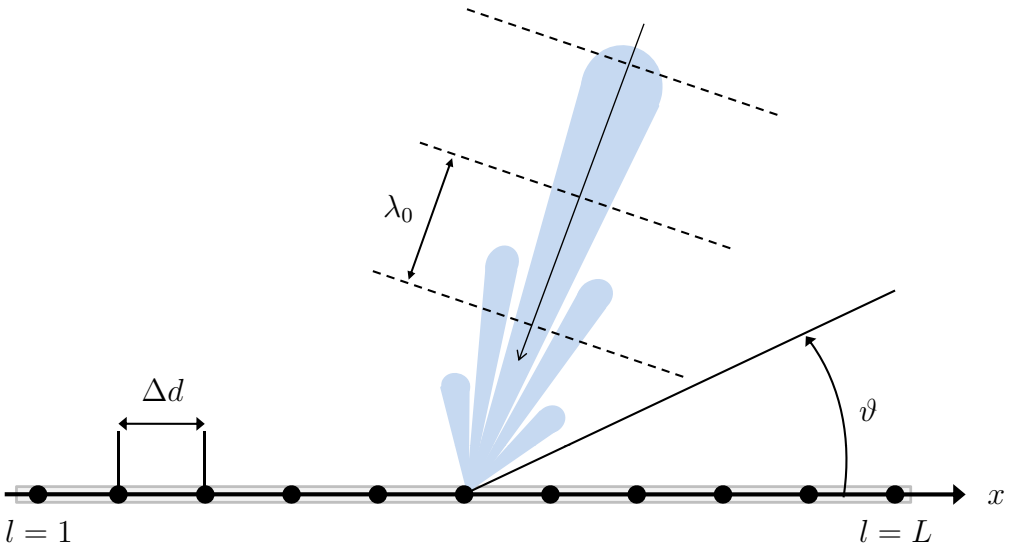



Figure 3.1. Direction finding for monochromatic plane-wave signals based on a uniform linear array.

## 3.2 Fiber-Optic Sensing

Fiber-optic sensors are versatile devices that can be used in a wide range of applications such as structural health monitoring [69, 70], chemical sensing [71], tomography [72] or medical applications [73]. As compared to electrical sensors, fiber-optic sensors benefit from different features of optical fibers, including low weight, broad bandwidth, low sensitivity to electro-magnetic interference and robustness to rough environmental conditions [74]. While various sensing technologies exist [75–81], *Fiber Bragg grating* (FBG) sensors have been extensively studied for high-resolution quasi-distributed temperature or strain monitoring [82–85]. They are particularly useful in 'smart structures' [75, 77–80] for an early detection of structural flaws, which can help to prevent accidents and to avoid high maintenance costs. Subsequently, FBG-based fiber optic sensing based on *wavelength-division multiplexing* (WDM) is considered. As compared to broadband light sources, wavelength-tunable lasers feature high local *signal-to-noise ratios* (SNRs) [70, 86].

The sensing principle and the signal model below are partly taken from [87][4], [88][5], [30][6].

### 3.2.1 Sensing Principle

An FBG describes a local variation of the refractive index [89]. FBGs are transparent to wavelengths except for a narrow region around the 'Bragg wavelength', $\lambda_B = 2n_{\text{eff}}\Lambda_{\text{FBG}}$ [90], where $n_{\text{eff}}$ is the effective refractive index of the propagating mode and $\Lambda_{\text{FBG}}$ is the grating period. If an FBG experiences a perturbation, typically strain or temperature variation [74], the grating period is either compressed or stretched. A uniform perturbation along the spatial extent of an FBG causes a total shift, $\delta\lambda_B$, of the Bragg wavelength [76, 91]. By measuring this shift, one is able to retrieve information about the perturbation acting at the FBG. An FBG-based fiber-optic sensor constitutes different FBGs that are imprinted along the fiber core to realize quasi-distributed sensing. In WDM-systems, their Bragg wavelengths are detuned, which enables an individual interrogation of the FBGs at different wavelengths, using either a broadband light source or a wavelength-tunable laser [86]. For an applied strain, $\epsilon_s$, or temperature change, $\Delta T$, the wavelength-shift can be quantified by [92–94]

$$\frac{\delta\lambda_B}{\lambda_B} = P_e\epsilon_s + [P_e(\alpha_s - \alpha_f) + \eta_t]\Delta T, \tag{3.6}$$

where $P_e$ is the strain-optic coefficient, $\alpha_s$, $\alpha_f$ are the thermal expansion coefficients of the fiber bonding material and the fiber itself, respectively. The variable $\eta_t$ denotes the thermo-optic coefficient. Moreover, the two types of perturbation can be unambiguously distinguished in the case where the temperature and strain coupling coefficients of two adjacent gratings are known [94].

### 3.2.2 Signal Model

In the scope of this thesis, a wavelength-tunable laser operating in pulsed mode is considered. Due to a limited bandwidth of the photo-detector and the receiver circuitry, individual short pulses are not resolved and only the pulse-envelope signal is available

[4]C. Weiss and A. M. Zoubir, "Fiber Sensing Using UFWT-Lasers and Sparse Acquisition," in *Proc. of the 21st European Signal Processing Conference (EUSIPCO)*, Marrakech, Morocco, Sep, 2013.

[5]C. Weiss and A. M. Zoubir, "Fiber Sensing Using Wavelength-Swept Lasers: A Compressed Sampling Approach," in *Proc. of the 3rd International Workshop on Compressed Sensing Theory and its Applications to Radar, Sonar and Remote Sensing (CoSeRa)*, Pisa, Italy, June, 2015.

[6]C. Weiss and A. M. Zoubir, "A Compressed Sampling and Dictionary Learning Framework for Wavelength-Division-Multiplexing-Based Distributed Fiber Sensing," accepted for publication in *Journal of the Optical Society of America A*, 2017 (assigned issue: vol. 34, no. 5).

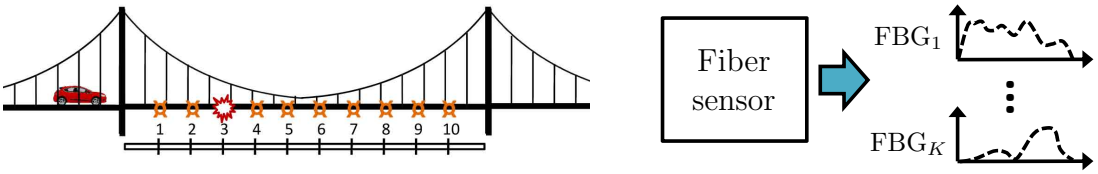



Figure 3.2. Schematic of a fiber sensing system for a suspension bridge: detection of strain and vibrations caused by crossing cars. Reference for the picture of the car: *Artist: spadassin (created: Nov. 15, 2014) title: "Car DS4" [online image], retrieved: Aug. 8, 2016 from https://openclipart.org/detail/202079/car-ds4.*

at the receiver. The perturbation profile is assumed to be stationary during the acquisition process. When the lasing wavelength is swept through the entire wavelength region of the gain medium, $R_{\mathrm{sw}}$, the signal is reflected if its spectral support overlaps with the narrow reflection region around the Bragg wavelength of FBG$_k$, $k = 1, \ldots, K$. It is assumed that, $\forall\ k = 1, \ldots, K$, the reflections are temporally sufficiently separated and can be unambiguously assigned to the individual FBGs. This assumption must be ensured by a proper system design. Further, the laser sweep rate, $S_{\mathrm{r}}$, is assumed to be constant, such that the nominal reflections in the unperturbed case are observed at the receiver at well-defined delays, $\tau_k$, $k = 1, \ldots, K$. One of the FBGs, say FBG$_j, j \in \{1, \ldots, K\}$, may act as a static reference point, which is located in a controlled environment. Then, the delays of the reflections from all other FBGs can be determined relative to this reference point. The differential delays are defined by

$$\Delta\tau_{jk} = \left|(\tau_k \pm \delta\tau_k) - \tau_j\right|, \quad k = 1, \ldots, K, \quad k \neq j. \tag{3.7}$$

where $\delta\tau_k$ denotes the absolute delay-shift due to impairments at FBG$_k$. In this setting, the wavelength-shift, $\delta\lambda_{B,k}$, can be inferred from the time-delays, $\delta\tau_k$ (or $\Delta\tau_{jk}$), by [70]

$$\delta\lambda_{B,k} = \delta\tau_k\, S_{\mathrm{r}}\, R_{\mathrm{sw}}, \quad k = 1, \ldots, K. \tag{3.8}$$

Finally, the amount and/or the nature of impairments can be inferred from (3.6).
The total received signal, $r(t)$, is described by the superposition of the reflections from all FBGs. Let $r_k(t)$, $k = 1, \ldots, K$, denote the reflections centered around zero. Then, the actual reflections can be modeled by vectors $\mathbf{r}_k$, that contain samples of the correctly delayed reflections, $r_k(t - \tau_k)$, $k = 1, \ldots, K$. When i.i.d. Gaussian measurement noise with variance $\sigma_n^2$ is assumed, the noisy observations become

$$\mathbf{y} = \sum_{i}^{K} \mathbf{r}_k + \mathbf{n}\,, \tag{3.9}$$

where $n_l \sim \mathcal{N}(0, \sigma_n^2)$, $l = 1, \ldots, L$. In order to obtain a sparse model for estimating the reflection delays, a dictionary can be assembled by concatenating different sub-dictionaries. Each sub-dictionary is associated with one of the FBGs and represents a

grid of possible delays for the reflection from that FBG. This yields an overall 'composite shift-invariant dictionary', which is created as follows:

For an initial time instant, $t_0$, the total observation time interval of one laser sweep, $I_{\text{ob}} = [t_0, t_0 + T_{\text{sw}}]$, is subdivided into $K$ non-overlapping segments. These segments are described by sets of delays, $\mathcal{T}_k,\ k = 1, \ldots, K$, where the $k$-th set represents the considered range of possible delays for the reflections from the $k$-th FBG. When the complete set of all dictionary indices is denoted by $\Omega = \{1, \ldots, N\}$, a subset

$$\Omega_k = \{n \in \Omega \,|\, N_{k-1} \leq n \leq N_k\}, \quad k = 1, \ldots, K, \tag{3.10}$$

can be associated with the reflections that fall inside the $k$-th segment of $I_{\text{ob}}$, where $N_{k+1} > N_k$ with $N_0 = 1$, $N_K = N = |\Omega|$. Then, also the dictionary, $\mathbf{D}$, can be split into $K$ sub-dictionaries, $\mathbf{D}_k,\ k = 1, \ldots, K$, where $\mathbf{D}_k$ corresponds to the respective index subset, $\Omega_k$. It is assumed that all sub-dictionaries contain the same number of atoms, i.e. $|\Omega_k| = (N/K) \in \mathbb{N}_+ \ \forall\ k = 1, \ldots, K$. The atoms of the $k$-th sub-dictionary can be formed by sampled and delayed versions of $r_k(t),\ k = 1, \ldots, K$, i.e.

$$[\mathbf{d}_n]_l = r_k(lT_d - n\delta t)\ , \qquad n \in \Omega_k, \quad l = 1, \ldots, L. \tag{3.11}$$

Herein, $T_d$ is the design sampling period. The granularity of the grid, hence, the dictionary coherence, is determined by the time increments, $\delta t > 0$, provided that $N\delta t \leq LT_d \leq T_{\text{sw}}$. Thus, a dictionary atom, $\mathbf{d}_n$, $n \in \Omega_k$, describes a reflection from the $k$-th FBG with a delay $\tau \in \mathcal{T}_k$, where

$$\mathcal{T}_k = \{\tau \in \mathbb{R}_+ \,|\, \tau = (n-1)\delta t,\ n \in \Omega_k\}. \tag{3.12}$$

When the sub-dictionaries are created in this way, $\mathbf{D}$ has a composite shift-invariant structure and is given by

$$\mathbf{D} = [\mathbf{D}_1, \ldots, \mathbf{D}_k, \ldots, \mathbf{D}_K]\ . \tag{3.13}$$

When $\mathbf{x}$ is a sparse (or compressible) representation of the total received signal, then the significant components in $\mathbf{x}$, i.e. $\mathcal{S}$, indicate the time delays $\tau_k \pm \delta\tau_k$ of the FBG reflections with respect to the dictionary in (3.13). Then, the sparse signal model can be written by

$$\mathbf{y} = \mathbf{D}\mathbf{x} + \mathbf{n}\,. \tag{3.14}$$

A detailed model for the reflections, $r_k(t),\ k = 1, \ldots, K$, is provided in Chapter 5 for a specific system architecture .

## 3.3 Compressed Sampling

*Compressed sensing* (CS) is an emerging technology that can lower the sampling rate below Nyquist, thereby reducing the number of samples and, in certain cases, even the hardware costs [20, 22, 26]. These properties make CS a favorable tool in various applications such as radar, tomography, remote sensing, imaging or coding [5, 16–18, 22, 25, 95–97]. The CS samples are acquired in a periodic non-uniform fashion [98, 99]. There exist digital and analog CS acquisition techniques. Digital CS is concerned with an efficient encoding of the original data samples [100], while analog CS focuses on the acquisition technique itself and how compression can be achieved in the analog domain before sampling [22, 101]. For analog CS, serial and parallel architectures have been developed. The inherent convolution operation in serial implementations produces some redundancy and yields correlated samples. In parallel architectures, a number of low-rate *analog-to-digital converters* (ADCs) are employed in parallel branches [22, 101–104]. When the sampling grid is uniform with a sampling period $T_s$ and a sampling rate $1/T_s$, then $M$ parallel ADCs can work at a reduced sampling rate $1/(MT_s)$ [102].
The task of sparse signal processing in the context of CS is to extract the desired information based on the compressed measurements in an automated and efficient manner, thereby exploiting the sparsity of the signal when represented by an appropriate dictionary. The goal is to find a sparse representation using the compressed measurements and the methods from sparse estimation theory [22, 25].

### 3.3.1 Signal Model

CS-based acquisition can be described by $M \ll L$ subsequent projections of the signal with sampling vectors $\{\boldsymbol{\phi}_m^\top\}_{m=1}^M$, that form the rows of a sampling matrix, $\boldsymbol{\Phi} \in \mathbb{R}^{M\times L}$. Then, the noisy compressed measurements (CS samples) are given by [20]

$$\mathbf{y} = \sum_{m=1}^{M} \boldsymbol{\phi}_m^\top \mathbf{r} + \mathbf{n} = \boldsymbol{\Phi}\mathbf{D}\mathbf{x} + \mathbf{n}\,, \tag{3.15}$$

where $\mathbf{B} = \boldsymbol{\Phi}\mathbf{D} \in \mathbb{R}^{M\times N}$ is referred to as the 'combined sensing matrix'. It can be regarded as a general redundant dictionary. The noise components, $n_m$, $m = 1, \ldots, M$, are often assumed to be i.i.d. zero-mean Gaussian distributed with variance $\sigma_n^2$. However, the assumptions on the noise statistics depend on the considered application. Since $M \ll L$, CS-based acquisition notably reduces the number of samples to be stored and processed. Besides using parallel ADCs, it is possible to reduce the average

sampling rate if $\mathbf{\Phi}$ is a sparse matrix, as it is done e.g. in [30][7].
For a successful sparse reconstruction of the signal, $B$ is required to show low atom coherence, especially when the reconstruction algorithm relies on local optimality criteria [29, 105]. Similarly, the performance depends on the sparsity level of $\mathbf{x}$. An explicit construction of a deterministic CS sampling matrix for which these criteria hold is intractable, since verifying the RIP is NP-hard [25]. Due to the favorable RIP conditions of sub-Gaussian random matrices (c.f. Chapter 2), they are frequently applied in CS [20, 25]. It is commonly assumed, that the rows of $\mathbf{\Phi}$ cannot sparsely represent the columns of $\mathbf{D}$, according to the incoherence property in [20]. However, this may no longer be true when the dictionary is redundant, due to the correlation between the columns of the resulting sensing matrix, $\mathbf{B}$ [27]. Successful reconstruction can only be guaranteed when $\mathbf{B}$ fulfills the RIP conditions in Chapter 2.
In the ensuing chapters, three different CS sampling matrices are considered. Their entries, $\phi_{ij}$, $i, j = 1, \ldots, N$, are i.i.d. random variables, drawn from one of the distributions below:

1. Gaussian distribution:
The entries are standard normally distributed random variables, i.e. $\phi_{ij} \sim \mathcal{N}(0, 1)$.

2. Rademacher distribution:
The entries are binary random variables, i.e. $\phi_{ij} \in \{1, -1\}$ with equal probabilities.

3. *Database-friendly* (DF) distribution [106]:
The entries take values $\phi_{ij} \in \{1, 0, -1\}$ with probabilities $\{1/6,\ 2/3,\ 1/6\}$, respectively.

[7] C. Weiss and A. M. Zoubir, "A Compressed Sampling and Dictionary Learning Framework for Wavelength-Division-Multiplexing-Based Distributed Fiber Sensing," accepted for publication in *Journal of the Optical Society of America A*, 2017 (assigned issue: vol. 34, no. 5).

# Chapter 4

# Regularization and Dictionary Calibration for Sparse Estimation

In this chapter, the problem of sparse regularization for $\ell_1$-mnimization in the presence of model errors is considered. Different techniques for sparse regularization as well as a dictionary calibration method are presented. High-resolution sparse estimation is performed using a redundant parametric dictionary with uncertainty. While different techniques can be used to take uncertainty in the dictionary into account, the model in this chapter considers an error term, $\Delta\mathbf{D}$, which is added to the ideal dictionary, $\mathbf{D}$, such that the true dictionary is given by $\mathbf{D} + \Delta\mathbf{D}$. Regarding the aspect of sparse estimation, it is assumed that the necessary RIP conditions for basic $\ell_1$-minimization are fulfilled. The general sparse estimation framework is specified for the application of direction finding in Section 3.1.

The material presented in this chapter is partly taken from [67][1], [66][2], [68][3].

### Chapter Outline

Section 4.1 gives an introduction and motivation to the considered problem, while Section 4.2 provides an overview of state-of-the-art techniques and related work. In Section 4.3, the problem statement and the considered probabilistic models for certain model errors are introduced. In Section 4.4 and 4.5, regularization techniques for sensor position errors and general phase/gain mismatches are presented. Next, in Section 4.6, a dictionary calibration method is shown. Some statements about the computational complexity are made in Section 4.7. Section 4.8 provides a discussion about the results and findings and Section 4.9 concludes gives a conclusion for this chapter.

[1] C. Weiss and A. M. Zoubir, "DOA Estimation in the Presence of Array Imperfections: A Sparse Regularization Parameter Selection Problem," in *Proc. of the IEEE Workshop on Statistical Signal Processing (SSP)*, Gold Coast, Australia, June/July, 2014.

[2] C. Weiss A. M. Zoubir, "A Sparse Regularization Technique for Source Localization with Non-uniform Sensor Gain," in *Proc. of the IEEE 8th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, A Coruña, Spain, June, 2014.

[3] C. Weiss and A. M. Zoubir, "Robust High-Resolution DOA Estimation with Array Pre-Calibration," in *Proc. of the 22nd European Signal Processing Conference (EUSIPCO)*, Lisbon, Portugal, September, 2014.

## 4.1 Introduction and Motivation

The performance of practical systems is often limited by different factors. Besides measurement noise, also model errors or uncertainty in the underlying signal model play an important role. Model errors and uncertainty may arise from a limited fabrication accuracy in the manufacturing process and are also affected by external factors such as temperature fluctuations or erosion. This leads to imperfect knowledge of particular system features such as the system geometry or the sensor characteristics. When the sparsity-promoting dictionary is created from a physical model with uncertainty, the achievable estimation performance is limited by the quality and correctness of the model. For sparse estimation based on $\ell_1$-minimization, the regularization parameter of the optimization problem is an essential component. It adjusts the tolerance level in order to achieve robustness and algorithmic stability. Therefore, a systematic strategy to automatically determine a suitable regularization parameter is desirable, that accounts for the joint effects of both model errors and noise. A statistical framework is adopted to obtain an upper bound of the *mean-squared error* (MSE), that is used for regularization. Herein, a direct relation between the physical system parameters and the regularization parameter is established. This method achieves robustness to model errors and a stable behavior of the sparse estimation algorithm. However, model errors lower the quality of the dictionary and, hence, the estimation accuracy.

Improved accuracy and robustness against model uncertainties can be achieved using dictionary calibration. While regularization techniques alone can only adjust the tolerance level to deal with certain types of model errors, dictionary calibration achieves robustness by jointly taking different types of model errors and noise into account, prior to sparse estimation. In contrast to creating the dictionary from an ideal model, a properly calibrated dictionary significantly alleviates the problem of choosing the regularization parameter. That is, the sensitivity of the sparse estimation algorithm to the regularization parameter is strongly reduced. In turn, when the regularization parameter is fixed, a calibrated dictionary yields robust sparse estimation performance for a wide range of SNRs.

The presented techniques require a parametric signal model in order to create a sparsity-promoting redundant dictionary and to incorporate application-specific model errors in terms of uncertain dictionary parameters. In this chapter, the general and abstract signal model for sparse estimation is specified for the application of direction finding, using the basic model in Chapter 3.1. Uncertainty is considered in the individual sensor gains and in the sensor positions. Dictionary calibration is performed by robust steering vector estimation, which yields a modified and calibrated dictionary. The performance of the presented techniques is shown for various scenarios of different SNRs and model errors in comparison to existing robust methods,

namely *Robust Capon Beamforming* (R-Capon) [107,108] and *Weighted Multiple Signal Classification* (W-MUSIC) [109]. The results obtained by a standard *delay-and-sum beamformer* (DSB) [64] serve as a non-robust reference.

## 4.2 State of the Art and Related Work

The problem of choosing the regularization parameter for sparse estimation is frequently addressed in the literature [5, 110–113]. In the development of an automated selection procedure, some authors exploit available knowledge of the noise statistics to determine a suitable regularization parameter. For the case of additive white Gaussian noise, a range of different regularization methods exists [5, 114, 115]. In [5], the discrepancy principle [116] is used to find a regularization parameter, such that the statistics of the residual error between the data and the signal model match the assumed noise statistics. In [114], a regularization parameter is obtained based on the square of the Frobenius norm of the noise component. A transformation of the noise components is performed in [115] to obtain an upper bound, that is used for regularization. In an alternative approach, Liu and Zhou [117] present a unified Bayesian framework for DOA estimation in the presence of array imperfections. When the noise statistics are unknown, the regularization parameter can be chosen based on the 'L-curve', which trades off the size of the regularized solution with the size of the corresponding residual [118,119]. In contrast to that, the presented regularization methods in this chapter assume knowledge of the noise statistics and represent an extension of the ideas in [5] by incorporating different model errors in addition to the measurement noise.
Some related techniques, that take uncertainty in the dictionary into account, have been reported: Chu *et al.* [120] address the problem of regularization under model uncertainties for acoustic imaging. The authors in [121] consider diagonal loading to reduce the sensitivity of the sparse solution to the regularization parameter. Additional constraints in the optimization problem are introduced by Shi *et al.* [122] in order to handle model uncertainties. The above methods, however, consider a general error term, that combines the contributions of several sources of uncertainty. Hence, they are not able to react to particular sources of uncertainty or model errors in a controlled way. The regularization methods presented in this chapter adopt a statistical framework to estimate the regularization parameter of the resulting optimization problem and establish a relationship to the physical parameters of interest, namely sensor position errors and general gain and phase mismatches. When *a priori* information is available, it can be effectively used to estimate a regularization parameter that compensates for specific impairments.

Rather than adjusting the tolerance level by choosing an appropriate regularization parameter, calibration techniques have been proposed to handle model errors and to determine the actual system parameters. For direction finding, a projection approach for array calibration is presented in [123]. Moreover, the work in [124] considers a recursive method to handle uncertainties in the array steering vectors for robust adaptive beamforming. The dictionary calibration technique presented in this chapter relies on these ideas to yield a modified and calibrated dictionary that incorporates model errors and reduces the sensitivity of the sparse estimation algorithm to the regularization parameter.

## 4.3 Problem Statement

In this chapter, the basic model for direction finding in Chapter 3.1 and in [5] is adopted to describe the incoming signals, the array geometry and the redundant dictionary, $\mathbf{D} \in \mathbb{C}^{L \times N}$, containing the $L$-dimensional (complex-valued) array steering vectors for $N$ specific angles. It is assumed that there exist $K$ point-like sources in the farfield, that emit monochromatic signals of wavelength $\lambda_0$ and circular-symmetric complex Gaussian distributed amplitudes with variance $\sigma_k^2$, $u_k(t) \sim \mathcal{CN}(0, \sigma_k^2)$, $k = 1, \ldots, K$, where $t$ is the time index. A number of $T$ snapshots are observed at time instances $t_\nu$, $\nu = 1, \ldots, T$, and collected in a matrix $\mathbf{Y} = [\mathbf{y}(t_1), \ldots, \mathbf{y}(t_T)]$. For each snapshot, a sparse vector, $\mathbf{x}(t_\nu)$, and a noise vector, $\mathbf{n}(t_\nu)$, are defined and collected in the matrices $\mathbf{X} = [\mathbf{x}(t_1), \ldots, \mathbf{x}(t_T)] \in \mathbb{C}^{N \times T}$ and $\mathbf{N} = [\mathbf{n}(t_1), \ldots, \mathbf{n}(t_T)] \in \mathbb{C}^{L \times T}$, respectively. The noise components are assumed to be zero-mean, circular-symmetric complex Gaussian distributed with variance $\sigma_n^2$, i.e. $n_l(t_\nu) \sim \mathcal{CN}(0, \sigma_n^2), l = 1, ..., L, \nu = 1, ..., T$. Given the dictionary, $\mathbf{D}$, the true DOAs of the signals are assumed to appear at the same dictionary indices for all snapshots. These indices are collected in the set of significant sparse coefficients, $\mathcal{S} = \{s_1, \ldots, s_K\}$, with cardinality $|\mathcal{S}| = K$. Ideally, the non-zero components in $\mathbf{x}(t_\nu)$ are equal to the signal amplitudes, i.e. $x_{s_k}(t_\nu) = u_k \; \forall \, \nu = 1, \ldots, T$, $k = 1, \ldots, K$. Model errors due to gain and phase mismatches are considered in terms of uncertainty in the dictionary. The true (but unknown) dictionary is denoted by

$$\tilde{\mathbf{D}} = \mathbf{D} + \Delta\mathbf{D}\,, \tag{4.1}$$

where $\mathbf{D}$ is the dictionary derived under ideal conditions and $\Delta\mathbf{D}$ is a correction term due to model errors. The extended signal model for $T$ snapshots, including uncertainty, is given by

$$\mathbf{Y} \;=\; \tilde{\mathbf{D}}\mathbf{X} + \mathbf{N} \;=\; (\mathbf{D} + \Delta\mathbf{D})\mathbf{X} + \mathbf{N}\,. \tag{4.2}$$

It is assumed that the measurement noise, $\mathbf{N}$, the source signals in $\mathbf{X}$, and the model errors, $\Delta\mathbf{D}$, are mutually independent. The goal is to estimate the DOAs of the incoming signals at high resolution by solving a particular instance of the $\ell_1$- minimization problem [5], i.e.

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}\in\mathcal{C}^N} \|\mathbf{x}^{(\ell_2)}\|_1 \qquad \text{s.t.} \qquad \|\mathbf{Y}-\mathbf{D}\mathbf{X}\|_2^2 \;\leq\; \beta_{\text{reg}}\,, \tag{4.3}$$

where $\forall\; n = 1,\ldots,N$ and $\forall\; \nu = 1,\ldots,T$:

$$x_n^{(\ell_2)} \;=\; \|\,[x_n(t_1),\ldots,x_n(t_T)]^\top\,\|_2\,, \tag{4.4}$$

$$|x_n(t_\nu)| \;=\; \sqrt{\mathsf{Re}\{x_n(t_\nu)\}^2 + \mathsf{Im}\{x_n(t_\nu)\}^2}\,. \tag{4.5}$$

Taking into account that the data is complex-valued, Equation (4.3) can be be rewritten as a second-order cone program [5]:

$$\begin{aligned}
&\min\; q_{\text{aux}} \\
\text{s.t.}\quad & \mathbf{1}^\top\mathbf{r}_{\text{aux}} \;\leq\; q_{\text{aux}}\,, \\
& \left\| \begin{array}{c} \mathbf{Y}-\mathbf{D}\mathbf{X} \\ (\beta_{\text{reg}}-1)/2 \end{array} \right\|_2 \;\leq\; (\beta_{\text{reg}}-1)/2\,, \\
& x_n^{(\ell_2)} \;\leq\; r_{n,\text{aux}}\,, \qquad\qquad n = 1,\ldots,N,
\end{aligned} \tag{4.6}$$

where $q_{\text{aux}}$ and $\mathbf{r}_{\text{aux}}$ are auxiliary parameters. Efficient solvers such as SeDuMi [125] can be used to solve Problem (4.6). However, this solution is sensitive to the regularization parameter, $\beta_{\text{reg}}$, especially if the observed data cannot be exactly described by the model due to model errors and uncertainty in the dictionary. Therefore, an appropriate sparse regularization method has to account for these effects. Later on, dictionary calibration is used to obtain a modified and calibrated dictionary, that better mimics the observed data in order to alleviate the regularization problem.

### 4.3.1 Probabilistic Models for Array Imperfections

Typical problems in practical systems are errors in the sensor gains and in the phase of the sensor signals. They may arise, for example, from inaccuracies in the fabrication process or external effects such as temperature fluctuations or erosion. Figure 4.1 shows a typical antenna pattern of an ideal and an impaired uniform linear array. It depicts the sensitivity of the array in dependence of the azimuth angle, $\vartheta$. Two particular types of model errors are considered: sensor position errors and general gain and phase errors. For each type, a probabilistic model is developed.

The presented models have been previously published in [67][4], [66][5], [68][6].

#### 4.3.1.1 Sensor Position Errors

Sensor position errors cause inaccuracies in the expected time delays, which leads to mismatches in the signal phase. The uniform linear array is assumed to be aligned with the $x$-axis and only errors in $x$-direction are considered. According to the model in Chapter 3.1, the ideal sensor spacing is given by $\Delta d = \lambda_0/2$. The position error, $\delta d$, should be smaller than half of the sensor spacing, i.e. $\delta d \leq \Delta d/2$. Using (3.2), the corresponding absolute phase error at the $l$-th sensor, $l = 1, \ldots, L$, for a signal with DOA $\vartheta$ is given by

$$\begin{aligned} |\Delta\overline{\phi}_l(\vartheta)| &= \left| \frac{\omega_0}{c_p}[\,(l-1)\Delta d + \delta d\,]\cos(\vartheta) \;-\; \frac{\omega_0}{c_p}[\,(l-1)\Delta d\,]\cos(\vartheta) \right| \\ &= \left| \frac{\omega_0}{c_p}\,\delta d\cos(\vartheta) \right| \quad \leq \quad \frac{\omega_0}{c_p}\,\frac{\Delta d}{2} \;=\; \frac{\omega_0}{c_p}\,\frac{\lambda_0}{4} \;=\; \frac{\pi}{2}. \end{aligned} \tag{4.7}$$

For simplicity, it is assumed that all phase errors, $\Delta\overline{\phi}_l(\vartheta)$, can be described by general i.i.d. zero-mean, Gaussian distributed variables with variance $\sigma_\phi^2$, i.e. $\Delta\overline{\phi} \sim \mathcal{N}(0, \sigma_\phi^2)$ for any DOA $\vartheta \in [0°, 180°]$ and any index $l = 1, \ldots, L$. Given the relation in (4.7), the standard deviation of the phase error can be defined by $\sigma_\phi = p_\phi\,\frac{\pi}{2}$, where $p_\phi$ is a parameter that controls the error.

#### 4.3.1.2 General Gain and Phase Errors

In order to model general gain and phase errors, a complex-valued gain variable can be introduced for each sensor. It is assumed that mutual coupling between the sensors can be neglected, and that the complex-valued gain errors can be modeled as i.i.d. zero-mean, circular symmetric complex Gaussian random variables with variance $\sigma_g^2$, i.e. $\tilde{g}_l \sim \mathcal{CN}(0, \sigma_g^2),\ l = 1, ..., L$. For each sensor, the ideal gain value in absence of phase errors is denoted by $g_0 \in \mathbb{R}_+$. Similar to the case of sensor position errors, the standard deviation can be defined by $\sigma_g = p_g\,\sigma_{g,\max}$, where it is assumed that $0 \leq p_g \leq 1$ and $\sigma_{g,\max} = g_0/2$.

[4]C. Weiss and A. M. Zoubir, "DOA Estimation in the Presence of Array Imperfections: A Sparse Regularization Parameter Selection Problem," in *Proc. of the IEEE Workshop on Statistical Signal Processing (SSP)*, Gold Coast, Australia, June/July, 2014.

[5]C. Weiss A. M. Zoubir, "A Sparse Regularization Technique for Source Localization with Non-uniform Sensor Gain," in *Proc. of the IEEE 8th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, A Coruña, Spain, June, 2014.

[6]C. Weiss and A. M. Zoubir, "Robust High-Resolution DOA Estimation with Array Pre-Calibration," in *Proc. of the 22nd European Signal Processing Conference (EUSIPCO)*, Lisbon, Portugal, September, 2014.

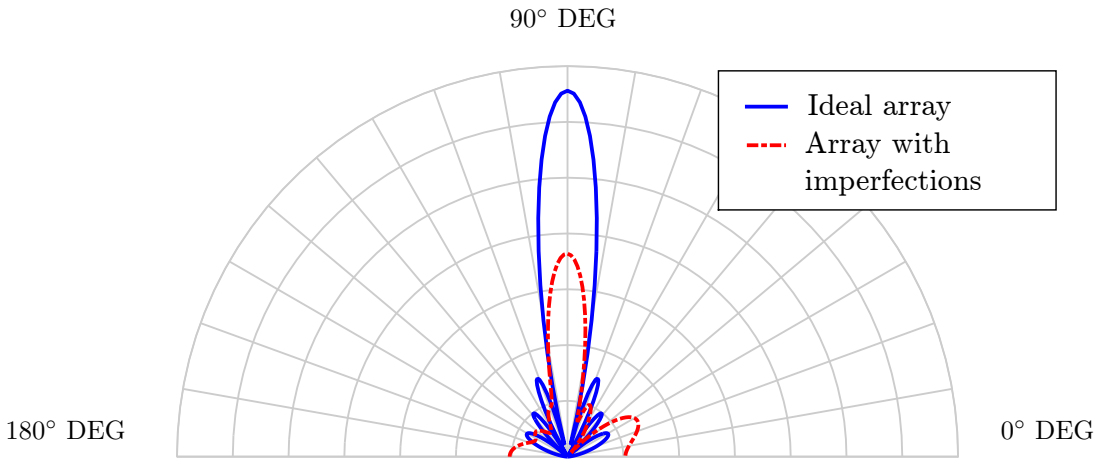



Figure 4.1. Antenna pattern of an ideal and impaired uniform linear array in dependence of the azimuth angle of the array, $\vartheta$. The range of interest lies within $[0, 180]^\circ$ DEG.

## 4.4 Sparse Regularization for Direction Finding With Sensor Position Errors

The material presented in this section is partly taken from [67][7].

Sensor position errors can be included in the general model in (4.2), where $\Delta\mathbf{D}$ contains the model errors (delay/phase errors) induced by imperfectly known sensor positions. The snapshots of the array become $\mathbf{Y} = (\tilde{\mathbf{D}}\mathbf{X} - \mathbf{D}\mathbf{X}) + \mathbf{D}\mathbf{X} + \mathbf{N}$. Following the ideas in [5], the regularization parameter, $\beta_{\text{reg}}$ in (4.3), can be derived from an upper bound for the *mean-squared error* (MSE) between the measured snapshots and those theoretically obtained under ideal model assumptions, i.e.

$$\mathbb{E}_{\mathbf{Y}} \, \|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_f^2 \;=\; \mathbb{E}_{\mathbf{Q},\mathbf{N}} \, \|\mathbf{Q} + \mathbf{N}\|_f^2 \,. \tag{4.8}$$

[7] C. Weiss and A. M. Zoubir, "DOA Estimation in the Presence of Array Imperfections: A Sparse Regularization Parameter Selection Problem," in *Proc. of the IEEE Workshop on Statistical Signal Processing (SSP)*, Gold Coast, Australia, June/July, 2014.

Herein, $\|\cdot\|_f$ signifies the Frobenius norm and $\mathbf{Q} = \Delta\mathbf{D}\,\mathbf{X} = [\mathbf{q}_1, \ldots, \mathbf{q}_T]^\top$ denotes the error term due to imperfectly known sensor positions, where $\mathbf{q}_\nu = (\tilde{\mathbf{D}} - \mathbf{D})\mathbf{x}(t_\nu)$, $\nu = 1 \ldots, T$. When the measurement noise is assumed to be independent of the model errors, 'Minkowski's inequality' [126] can be applied to obtain

$$\mathbb{E}_{\mathbf{Y}}\,\|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_f^2 \;\leq\; (\mathbb{E}_{\mathbf{Q}}\,\|\mathbf{Q}\|_f^2) \;+\; (\mathbb{E}_{\mathbf{N}}\,\|\mathbf{N}\|_f^2) \;+\; 2\,(\mathbb{E}_{\mathbf{N}}\,\|\mathbf{N}\|_f)\,(\mathbb{E}_{\mathbf{Q}}\,\|\mathbf{Q}\|_f). \quad (4.9)$$

In order to find an expression for this upper bound, the single terms on the right-hand side in (4.9) have to be evaluated.
For $\mathbb{E}_{\mathbf{N}}\,\|\mathbf{N}\|_f^2$, the authors in [5] obtain an upper limit by taking into account that every entry in $\mathbf{N}$ is zero-mean, circular symmetric complex Gaussian distributed. In particular [5],

$$\frac{[\mathbf{n}_\nu]_l}{\sigma_n} \;\sim\; \mathcal{CN}(0,1), \qquad l = 1, \ldots, L, \quad \nu = 1, \ldots, T, \quad (4.10)$$

with $\mathsf{Re}\{[\mathbf{n}_\nu]_l\} \sim \mathcal{N}(0, \sigma_n^2/2)$ and $\mathsf{Im}\{[\mathbf{n}_\nu]_l\} \sim \mathcal{N}(0, \sigma_n^2/2)$. Then, the total noise power over all snapshots is Chi-square distributed with $2TL$ degrees of freedom [5], i.e.

$$\frac{2}{\sigma_n^2}\|\mathbf{N}\|_f^2 \;\sim\; \mathcal{X}^2_{2TL}. \quad (4.11)$$

The noise term can be confined within a confidence interval for the $\mathcal{X}^2_{2TL}$-distribution with confidence level $\alpha_\chi$ [5],

$$\mathcal{X}^2_{l,\alpha_\chi} \;\leq\; \frac{2}{\sigma_n^2}\|\mathbf{N}\|_f^2 \;\leq\; \mathcal{X}^2_{u,\alpha_\chi}, \quad (4.12)$$

where $\mathcal{X}^2_{u,\alpha_\chi}$ and $\mathcal{X}^2_{l,\alpha_\chi}$ are the upper and lower bounds, respectively. An upper bound for the noise variance can be found by using (4.12) [5], i.e.

$$\sigma_n^2 \;\leq\; 2\,\frac{\|\mathbf{N}\|_f^2}{\mathcal{X}^2_{l,\alpha_\chi}}. \quad (4.13)$$

Ultimately, an upper bound for the noise term is found by using (4.12) [5], i.e.

$$\mathbb{E}_{\mathbf{N}}\,\|\mathbf{N}\|_f^2 \;\leq\; \mathcal{X}^2_{u,\alpha_\chi}\,\frac{\|\mathbf{N}\|_f^2}{\mathcal{X}^2_{l,\alpha_\chi}} \;\approx\; TL\,\mathcal{X}^2_{u,\alpha_\chi}\,\frac{\hat{\sigma}_n^2}{\mathcal{X}^2_{l,\alpha_\chi}} \;\triangleq\; \hat{n}_{\mathrm{MSE}}\,, \quad (4.14)$$

where $\hat{\sigma}_n^2$ is the estimated noise power. The confidence level, $\alpha_\chi$, controls the tightness of this upper bound. For a given level, $\alpha_\chi \cdot 100\%$ of all noise realizations have a smaller power than the upper bound in (4.14).

For the term $\mathbb{E}_{\mathbf{N}}\,\|\mathbf{N}\|_f$ in (4.9), a similar strategy is pursued. Since the square root of the total noise power is Chi-distributed with $2TL$ degrees of freedom, i.e.

$$\sqrt{2}\,\frac{\|\mathbf{N}\|_f}{\sigma_n} = \sqrt{\sum_{l=1}^{L}\sum_{\nu=1}^{T} 2\left(\frac{[\mathbf{n}_\nu]_l}{\sigma_n}\right)^2} \sim \mathcal{X}_{2TL}\,. \tag{4.15}$$

Given a confidence level, $\tilde{\alpha}_\chi$, the corresponding lower and upper confidence bounds, $\mathcal{X}_{l,\tilde{\alpha}_\chi}$ and $\mathcal{X}_{u,\tilde{\alpha}_\chi}$, respectively, can be obtained in the same way as above, by

$$\mathbb{E}_{\mathbf{N}}\,\|\mathbf{N}\|_f \;\leq\; \mathcal{X}_{u,\tilde{\alpha}_\chi}\frac{\|\mathbf{N}\|_f^2}{\mathcal{X}^2_{l,\tilde{\alpha}_\chi}} \;\approx\; TL\,\mathcal{X}_{u,\tilde{\alpha}_\chi}\frac{\hat{\sigma}_n^2}{\mathcal{X}^2_{l,\tilde{\alpha}_\chi}} \;\triangleq\; \hat{n}_{\mathrm{RMSE}}\,. \tag{4.16}$$

Next, an upper bound for $\mathbb{E}_{\mathbf{Q}}\,\|\mathbf{Q}\|_f^2$ is derived. For all snapshots, $\nu = 1,...,T$, the elements $\mathbf{q}_\nu$ are functions of the columns of $(\tilde{\mathbf{D}} - \mathbf{D})$, corresponding to the $K$ non-zero entries in $\mathbf{x}(t_\nu)$. Let $\tau_l(\vartheta_k)$ denote the ideal signal delays at all sensors and for all sources, and let $\overline{\phi}_l(\vartheta_k) = \omega_0\tau_l(\vartheta_k)$ be the corresponding phase terms $\forall\; k \in \mathcal{S}$, $l = 1,\ldots,L$. Then, the $l$-th entry in $\mathbf{q}_\nu,\; \nu = 1,\ldots,T$, can be written by

$$\begin{aligned}
[\mathbf{q}_\nu]_l &= \sum_{k\in\mathcal{S}} x_k(t_\nu)\,\left(\,\exp\!\big(-j[\,\overline{\phi}_l(\vartheta_k) + \Delta\overline{\phi}_{l,k}\,]\big) - \exp\!\big(-j\overline{\phi}_l(\vartheta_k)\big)\,\right)\\
&= \sum_{k\in\mathcal{S}} x_k(t_\nu)\,\exp\!\left(-j\left[\overline{\phi}_l(\vartheta_k) + \frac{\Delta\overline{\phi}_{l,k}}{2} + \frac{\pi}{2}\right]\right) 2\sin\left(\frac{\Delta\overline{\phi}_{l,k}}{2}\right).
\end{aligned} \tag{4.17}$$

The error model in Section 4.3 assumes that $\Delta\overline{\phi}_{l,k}$, $l = 1,\ldots,L$, $k \in \mathcal{S}$ are i.i.d. random variables and that $\mathbf{x}(t_\nu)$ and $\Delta\overline{\phi}_{l,k}$ are mutually independent $\forall\; l,k$, and $\forall\; \nu = 1,\ldots,T$. Hence, using the abbreviation $\psi_{l,k} = \overline{\phi}_l(\vartheta_k) + \frac{\Delta\overline{\phi}_{l,k}}{2}$, $l = 1,\ldots,L,\; k \in \mathcal{S}$, the $(\nu,r)$-th entry of $\mathbb{E}_{\mathbf{Q}}\;\mathbf{Q}^{\mathrm{H}}\mathbf{Q}$ becomes

$$\begin{aligned}
\mathbb{E}_{\mathbf{Q}}\;\mathbf{q}_v^{\mathrm{H}}\mathbf{q}_r &= \sum_{l=1}^{L}\sum_{k\in\mathcal{S}}\sum_{m\in\mathcal{S}}\big(\;\mathbb{E}_{\mathbf{x}}\;x_k^*(t_v)x_m(t_r)\;\big)\\
&\qquad\times\;\mathbb{E}_{\Delta\overline{\phi}_e}\,\exp(-j[\psi_{l,k} - \psi_{l,m}])\,4\sin\left(\frac{\Delta\overline{\phi}_{l,k}}{2}\right)\sin\left(\frac{\Delta\overline{\phi}_{l,m}}{2}\right).
\end{aligned} \tag{4.18}$$

Since the source signals are assumed to be zero-mean Gaussian variables with variance $\sigma_u^2$, one obtains $\mathbb{E}_{\mathbf{x}}\;x_k^*(t_\nu)x_m(t_r) = \sigma_u^2\delta_{km}\delta_{\nu r}$, where $\delta_{a,b},\;\; a,b \in \mathbb{N}$, denotes Kronecker's delta function.

Therefore, all elements with $k \neq m$ and $\nu \neq r$ can be omitted and (4.18) can be simplified as

$$\begin{aligned}\mathbb{E}_{\mathbf{Q}} \, \|\mathbf{Q}\|_f^2 &= \sigma_u^2 \, \mathbb{E}_{\Delta\overline{\phi}} \sum_{\nu=1}^{T} \sum_{l=1}^{L} \sum_{k \in \mathcal{S}} 4 \sin^2(\Delta\overline{\phi}_{l,k}) \\ &= \sigma_u^2 \, \mathbb{E}_{\Delta\overline{\phi}} \sum_{\nu=1}^{T} \sum_{l=1}^{L} \sum_{k \in \mathcal{S}} 2(1 - \cos(\Delta\overline{\phi}_{l,k})). \end{aligned} \tag{4.19}$$

Subsequently, the variables $\Delta\overline{\phi}_{l,k}, l = 1, \ldots, L$, $k \in \mathcal{S}$, are addressed using a canonical index $z$, i.e. $\Delta\overline{\phi}_z$, $z = 1, \ldots, LK$. Further, a variable transformation, $g_{\text{tr}}(\cdot)$, yields $q_\phi = g_{\text{tr}}(\Delta\overline{\phi}) = 2(1 - \cos(\Delta\overline{\phi}))$. The probability density $p_{q_\phi}$ can be obtained from the density $p_{\Delta\overline{\phi}}$ by [127]

$$\begin{aligned} p_{q_\phi}(q_\phi) &= p_{\Delta\overline{\phi}}\left( g_{\text{tr}}^{-1}(q_\phi) \right) \left| g_{\text{tr}}^{-1}(q_\phi) \right| \\ &= \frac{C_{q_\phi}^{-1}}{\sqrt{8\pi}\sigma_\phi} \exp\left( \frac{-[\cos^{-1}\left(1 - \frac{q_\phi}{2}\right)]^2}{2\sigma_\phi^2} \right) \left( 1 - \left( 1 - \frac{q_\phi}{2} \right)^2 \right)^{-\frac{1}{2}}, \end{aligned} \tag{4.20}$$

where $C_{q_\phi}$ is a normalization constant. The mean and the variance of $q_\phi$ are denoted by $\mu_{q_\phi}$ and $\sigma_{q_\phi}^2$, respectively. Using (4.19), one finds

$$\mathbb{E}_{\mathbf{Q}} \, \|\mathbf{Q}\|_f^2 = \sigma_u^2 \, \mathbb{E}_{\mathbf{q}_\phi} \sum_{\nu=1}^{T} \sum_{z=1}^{LK} q_{\phi,z} = T\sigma_u^2 \, (\mathbb{E}_{\tilde{q}_\phi} \, \tilde{q}_\phi) \,, \tag{4.21}$$

where $\tilde{q}_\phi = \sum_{z=1}^{LK} q_{\phi,z}$. By taking the limit $LK \to \infty$, the 'central-limit theorem' can be applied. Hence, $\tilde{q}_\phi \sim \mathcal{N}(LK\mu_{q_\phi}, LK\sigma_{q_\phi}^2)$. An upper bound for $\mathbb{E}_{\mathbf{Q}} \, \|\mathbf{Q}\|_f^2$ can be obtained in a similar fashion as above in (4.14) and (4.16), using a confidence interval for $\tilde{q}_\phi$ with confidence level $\alpha_\phi$ and corresponding upper bound $\mathcal{N}_{u,\alpha_\phi}$:

$$\mathbb{E}_{\mathbf{Q}} \, \|\mathbf{Q}\|_f^2 \leq T \, \hat{\sigma}_u^2 \mathcal{N}_{u,\alpha_\phi} \triangleq \hat{q}_{\text{MSE}} \,, \tag{4.22}$$

where $\hat{\sigma}_u^2$ is an estimate of $\sigma_u^2$.
Next, an upper bound for $\mathbb{E}_{\mathbf{Q}} \, \|\mathbf{Q}\|_f = \mathbb{E}_{\mathbf{Q}} \sqrt{\|\mathbf{Q}\|_f^2}$ is sought for. Keeping in mind that the square-root is a concave function, 'Jensen's inequality' [126] can be applied:

$$\mathbb{E}_{\mathbf{Q}} \sqrt{\|\mathbf{Q}\|_f^2} \leq \sqrt{\mathbb{E}_{\mathbf{Q}} \, \|\mathbf{Q}\|_f^2} \leq \sqrt{\hat{q}_{\text{MSE}}} \,. \tag{4.23}$$

Ultimately, using (4.14), (4.16), (4.22) and (4.23), an estimate for the regularization parameter in (4.9) is obtained by

$$\hat{\beta}_{\text{reg}} = \hat{q}_{\text{MSE}} + \hat{n}_{\text{MSE}} + 2 \, \hat{n}_{\text{RMSE}} \sqrt{\hat{q}_{\text{MSE}}} \,. \tag{4.24}$$

### 4.4.1 Simulations

In subsequent simulations, the sparse estimation technique is referred to as 'SPARSE' and the regularization parameter, $\beta_{\text{reg}}$, is estimated according to Section 4.4. SPARSE is compared to existing robust DOA estimation techniques, namely *Robust Capon Beamforming* (R-Capon) [107, 108] and *Weighted MUSIC* (W-MUSIC) [109]. Also the results of a standard *delay-and-sum beamformer* (DSB) [64] are shown and serve as a non-robust reference method. In all simulations, the considered angular range is rotated by $-90°$, such that $\vartheta \in [-90°, 90°]$.

#### 4.4.1.1 Setup

The simulations are conducted using a uniform linear array with $L = 9$ sensors and $K = 2$ sources with equal power. The dictionary contains $N = 181$ atoms corresponding to 181 equidistant angles in the range $[-90°, 90°]$ with an accuracy of $\delta\vartheta = 1°$. The sources are located at angles $\vartheta_1 = -45°$ and $\vartheta_2 = -35°$. This corresponds to a spacing of $10°$ and falls inside the Rayleigh resolution limit, which defines a minimally resolvable source separation [64] of

$$\Delta\vartheta_{\text{Rayl}} \;=\; \frac{\lambda_0}{L\,\Delta d}\,. \tag{4.25}$$

The correlation between the complex amplitudes of the source signals, $u_k$, with powers $\sigma_k,\ k = 1, 2$, is indicated by a correlation coefficient, $\zeta_c,\ 0 \leq \zeta_c \leq 1$. For zero-mean amplitudes, it is defined by

$$\zeta_c = \frac{\mathbb{E}\ (u_1\, u_2)}{\sigma_1\, \sigma_2}. \tag{4.26}$$

R-Capon and W-MUSIC require an estimate of the covariance matrix of the sensor measurements. Therefore, $T = 30$ snapshots are used for all methods. It is assumed that the number of sources, $K$, and the source power, $\sigma_u^2$, are known. In order to determine the regularization parameter, $\hat{\beta}_{\text{reg}}$, it is required to estimate the noise power, $\sigma_n^2$. In the presented simulations, it is assumed that pure noise samples are available for this purpose. The confidence levels are equally chosen as $\alpha_\chi = \tilde{\alpha}_\chi = \alpha_\phi = 0.9$. The results are shown for different SNRs and different values of $p_\phi$ and $\zeta_c$. The mean and the variance of $q_\phi$, i.e. $\mu_{q_\phi}$ and $\sigma_{q_\phi}^2$, are calculated numerically from (4.20). Empirically, it can be observed that the probability density of $\tilde{q}_\phi$ has an almost-Gaussian shape when $LK > 10$, which shows that the central limit theorem applied in (4.21) is reasonable for the considered scenarios.

#### 4.4.1.2 DOA Spectra

Figure 4.2 shows some typical realizations of the DOA spectra for each method. Figures 4.2(a)-4.2(f) show the results for SNR= 15 dB. Without correlation ($\zeta_c = 0$),

W-MUSIC, R-Capon and SPARSE are able to resolve both sources in the presence of sensor position errors up to $p_\phi = 0.65$. The best performance is achieved by SPARSE. Using the presented regularization technique, it can even handle stronger errors up to $p_\phi = 0.85$. However, without proper regularization, spurious peaks appear in the spectrum, which falsely indicate additional sources as shown in Figure 4.2(c). Figures 4.2(e)-4.2(f) show the impact of correlation on W-MUSIC and R-Capon. Since SPARSE does not make any assumptions on correlation, it is still able to resolve both sources even for $\zeta_c = 0.9$, which is depicted in Figure 4.2(e). Nevertheless, source correlation degrades the robustness of SPARSE to sensor position errors. In Figure 4.2(f) it is shown that, for $\zeta_c = 0.9$, SPARSE can no longer resolve the sources in the presence of smaller position errors with $p_\phi = 0.3$. Finally, Figures 4.2(g)-4.2(i) show the results for SNR= 0 dB. Although both sources can be resolved by W-MUSIC, R-Capon and SPARSE for $p_\phi = 0.3$, the accuracy of W-MUSIC and R-Capon is reduced, while SPARSE is only slightly affected. However, for $p_\phi = 0.65$ in Figure 4.2(i), none of the methods is able to resolve the two sources.

#### 4.4.1.3 Performance Evaluation

The performance is evaluated using 500 Monte Carlo trials in Figure 4.3. The upper parts of each figure show the angular error (in degrees) with respect to the grid accuracy, $\delta\vartheta = 1°$. The error is calculated based on the positions (indices) of the significant elements in $\mathbf{x}$, i.e the elements with the largest modulus contained in $\hat{\mathcal{S}}$. In particular, the estimates in $\hat{\mathcal{S}}$ with the smallest distance to the true source locations in $\mathcal{S}$ are used to calculate the error. The lower parts of each figure show the success rates. In a successful trial, the regularization is appropriately chosen. That is, the spectrum exhibits exactly two peaks, corresponding to the two simulated sources (note that the number of sources is assumed to be knwon for this evaluation). Figures 4.3(a)-4.3(c) show different results for uncorrelated signals, i.e. $\zeta_c = 0$. The success rate for SPARSE is always superior to that of W-MUSIC and R-Capon. When the success rate is high, the angular error is close to the grid accuracy for all methods. At high SNRs, SPARSE achieves high success rates for sensor position errors up to $p_\phi = 0.65$ (Figures 4.3(a)-4.3(c)). At low SNRs and $p_\phi = 0.65$ (Figure 4.3(c)), W-MUSIC and R-Capon yield higher success rates. Figures 4.3(d)-4.3(f) show the impact of correlation. W-MUSIC and R-Capon, are no longer able to achieve good success rates even for low values of $p_\phi$. Also SPARSE is affected by the joint impacts of noise, correlation, and sensor position errors. In Figure 4.3(e), W-MUSIC achieves slightly better success rates and lower errors than SPARSE for $\zeta_c = 0.5$ and $p_\phi = 0.3$. In the absence of model errors, SPARSE can handle moderate correlation levels up to $\zeta_c = 0.5$ (Figure 4.3(d)), but strong correlation levels reduce the success rate especially at low SNRs (Figure 4.3(f)). At low success rates, less Monte

Carlo trials are available, since the error is only calculated in case of a successfull trial. Therefore stronger fluctuations in the RMSE can be observed.

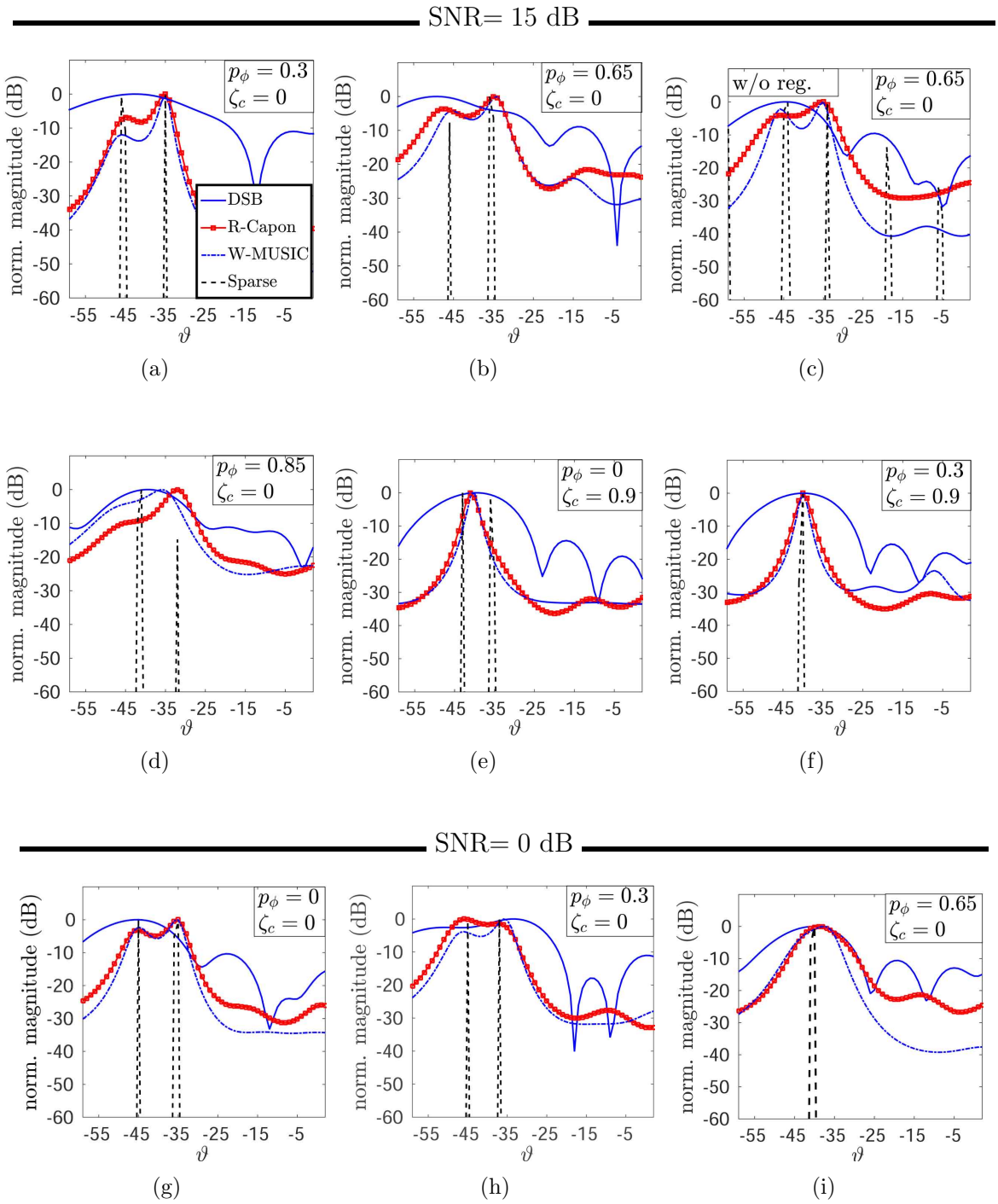



Figure 4.2. Comparison of the DOA spectra obtained by SPARSE, W-MUSIC and R-Capon. The results of a DSB are shown as a reference. For SPARSE, $\hat{\beta}$ was estimated according to Section 4.4.

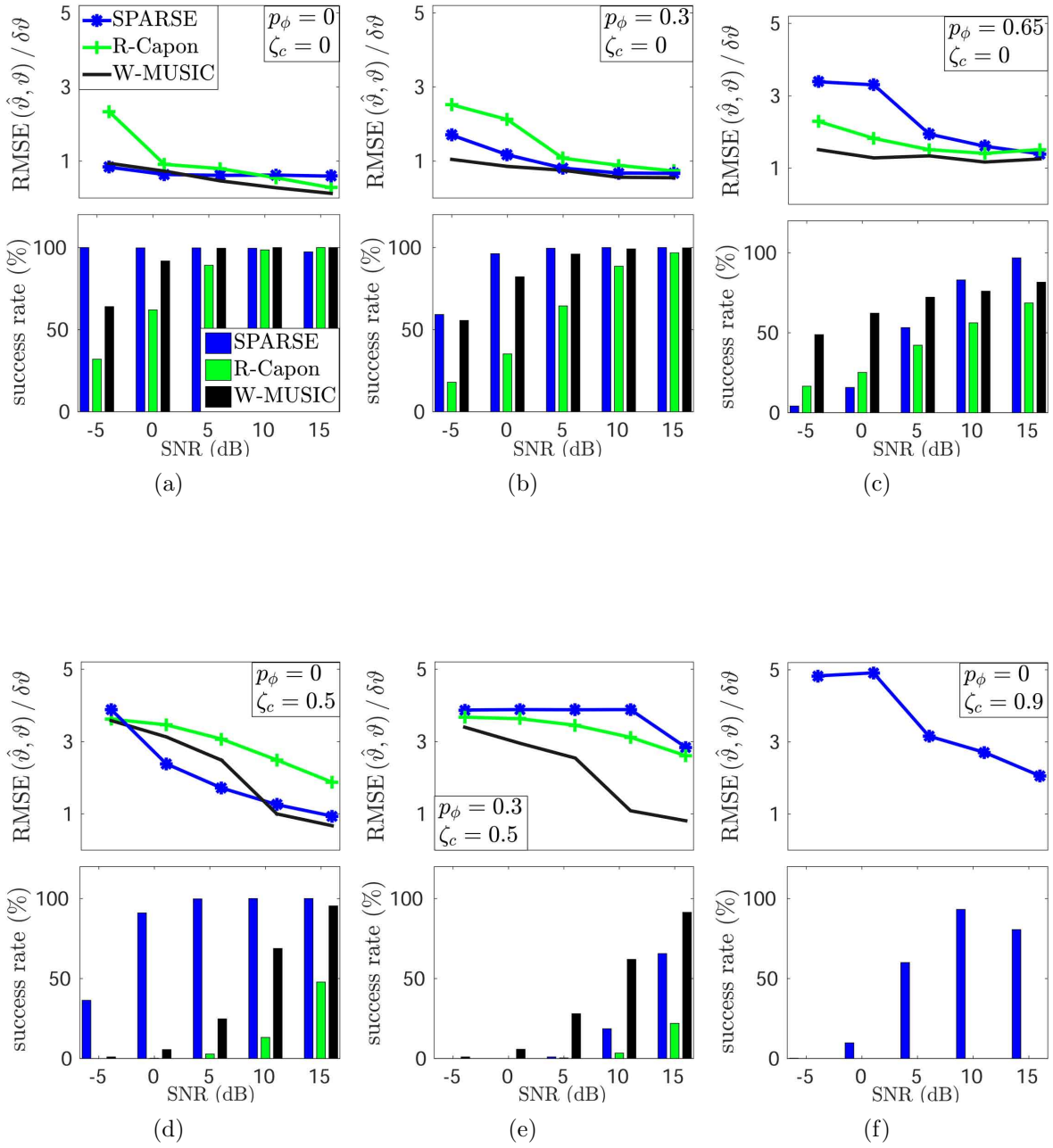



Figure 4.3. Performance of SPARSE, W-MUSIC and R-Capon: success rates (lower subfigures) and angular errors relative to the grid accuracy, $\delta\vartheta$, (upper subfigures) vs. different SNRs. For SPARSE, $\hat{\beta}$ was estimated according to Section 4.4.

## 4.5 Sparse Regularization for Direction Finding With Gain/Phase Errors

The material presented in this section is partly taken from [66][8].

When gain and phase errors are considered, the general model in (4.2) can be written by

$$\mathbf{Y} = (\mathbf{G}_0 + \tilde{\mathbf{G}})\mathbf{A}\mathbf{X} + \mathbf{N} \;, \tag{4.27}$$

where $\tilde{\mathbf{D}} = (\mathbf{G}_0 + \tilde{\mathbf{G}})\mathbf{A}$, $\mathbf{D} = \mathbf{G}_0\mathbf{A}$ and $\Delta\mathbf{D} = \tilde{\mathbf{G}}\mathbf{A}$. The matrix $\mathbf{A}$ contains the ideal steering vectors, $\mathbf{G}_0$ is the ideal sensor gain matrix and $\tilde{\mathbf{G}}$ models gain and phase errors. It is assumed that the sensors are not coupled, such that $\mathbf{G}_0 = g_0\mathbf{I}$ and $\tilde{\mathbf{G}} = \mathrm{diag}(\tilde{g}_1, \ldots, \tilde{g}_L)$, where $g_0 \in \mathbb{R}_+$ and $\tilde{g}_l \sim \mathcal{CN}(0, \sigma_g^2)$, $l = 1, \ldots, L$, according to Section 4.3.1.2. When $\tilde{\mathbf{G}} = \mathbf{0}$, the general $\ell_1$-minimization problem in (4.3) becomes

$$\min \|\mathbf{x}^{(\ell_2)}\|_1 \qquad \text{s.t.} \qquad \|\mathbf{Y} - g_0\mathbf{A}\mathbf{X}\|_f^2 \;\leq\; \beta_{\text{reg}} \;. \tag{4.28}$$

Herein, the hyperparameter $\beta_{\text{reg}}$ has to account for the joint effects of gain/phase errors and noise. Similar to the case of non-ideal sensor positions, $\beta_{\text{reg}}$ is estimated as an upper bound of the MSE between the true sensor measurements and those theoretically obtained under ideal model assumptions, i.e.

$$\mathbb{E}_{\mathbf{Y}} \; \|\mathbf{Y} - g_0\mathbf{A}\mathbf{X}\|_f^2 \;=\; \mathbb{E}_{\mathbf{Q},\mathbf{N}} \; \|\mathbf{Q} + \mathbf{N}\|_f^2 \tag{4.29}$$

where $\mathbf{Q} = [\mathbf{q}_1, .., \mathbf{q}_T] = \tilde{\mathbf{G}}\mathbf{A}\mathbf{X}$. The error model in Section 4.3 assumes that the measurement noise and the model errors are mutually independent. Then, using 'Minkowski's inequality' [126], an upper bound for the MSE can be obtained by

$$\mathbb{E}_{\mathbf{Q},\mathbf{N}} \; \|\mathbf{Q}+\mathbf{N}\|_f^2 \;\leq\; \left(\mathbb{E}_{\mathbf{Q}} \; \|\mathbf{Q}\|_f^2\right) \;+\; \left(\mathbb{E}_{\mathbf{N}} \; \|\mathbf{N}\|_f^2\right) \;+\; 2 \; \left(\mathbb{E}_{\mathbf{Q}} \; \|\mathbf{Q}\|_f\right) \left(\mathbb{E}_{\mathbf{N}} \; \|\mathbf{N}\|_f\right). \tag{4.30}$$

In order to obtain a useful expression for the right-hand side in (4.30), upper bounds for the individual expected values are subsequently derived.
Following the ideas in [5], the terms $\mathbb{E}_{\mathbf{N}} \; \|\mathbf{N}\|_f^2$ and $\mathbb{E}_{\mathbf{N}} \; \|\mathbf{N}\|_f$ are evaluated as in Section 4.4, using Equations (4.16) and (4.14). As in Section 4.4, the confidence levels of the $\mathcal{X}^2_{2TL}$ and the $\mathcal{X}_{2TL}$ distributions are denoted by $\alpha_\chi$ and $\tilde{\alpha}_\chi$, respectively. Also, the corresponding upper and lower bounds are equivalently defined by $\mathcal{X}^2_{u,\alpha_\chi}, \mathcal{X}^2_{l,\alpha_\chi}$ and $\mathcal{X}_{u,\tilde{\alpha}_\chi}, \mathcal{X}_{l,\tilde{\alpha}_\chi}$, respectively. Further, $\hat{\sigma}_n^2$ represents an estimate for $\sigma_n^2$.

[8] C. Weiss A. M. Zoubir, "A Sparse Regularization Technique for Source Localization with Non-uniform Sensor Gain," in *Proc. of the IEEE 8th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, A Coruña, Spain, June, 2014.

Regarding the term $\mathbb{E}_{\mathbf{Q}} \, \|\mathbf{Q}\|_f^2$ in (4.30), the individual entries in $\mathbf{Q}$ can be written by

$$[\mathbf{q}_\nu]_l = \sum_{k \in \mathcal{S}} x_k(t_\nu) \, \tilde{g}_l \exp\big( -j \overline{\phi}_l(\vartheta_k) \big), \qquad \nu = 1, ..., T, \; l = 1, \ldots, L. \tag{4.31}$$

It is assumed that the entries in $\mathbf{X}$ and $\tilde{\mathbf{G}}$ are independent. Then, $\forall \; \nu, r = 1, .., T$, the $(\nu, r)$-th entry of $\mathbb{E}_{\mathbf{Q}} \; \mathbf{Q}^{\mathrm{H}}\mathbf{Q}$ becomes

$$\mathbb{E}_{\mathbf{Q}} \; \mathbf{q}_\nu^{\mathrm{H}} \mathbf{q}_r = \left( \mathbb{E}_{\tilde{\mathbf{G}}} \sum_{l=1}^{L} \tilde{g}_l^* \tilde{g}_l \right) \sum_{k \in \mathcal{S}} \sum_{m \in \mathcal{S}} \exp\big( j [\, \overline{\phi}_l(\vartheta_m) - \overline{\phi}_l(\vartheta_k) \,] \big) \; (\mathbb{E}_{\mathbf{x}} \; x_k^*(t_\nu) x_m(t_r)). \tag{4.32}$$

Since the amplitudes of the source signals are assumed to be zero-mean Gaussian variables with variance $\sigma_u^2$, one obtains $\mathbb{E}_{\mathbf{x}} \; x_k^*(t_\nu) x_m(t_r) = \sigma_u^2 \, \delta_{km} \, \delta_{\nu r}$, where $\sigma_u^2$ can be replaced by an estimate of the source power, $\hat{\sigma}_u^2$. Herein, $\delta_{ab}, \; a, b \in \mathbb{N}$, denotes 'Kronecker's delta function'. Thus, all terms with $k \neq m$ or $\nu \neq r$ vanish, which yields

$$\mathbb{E}_{\mathbf{Q}} \; \|\mathbf{Q}\|_f^2 \;\; = \;\; \sum_{\nu=1}^{T} \left( \mathbb{E}_{\mathbf{Q}} \; \mathbf{q}_\nu^{\mathrm{H}} \mathbf{q}_\nu \right) \;\; \approx \;\; TK \hat{\sigma}_u^2 \left( \mathbb{E}_{\tilde{\mathbf{G}}} \; \sum_{l=1}^{L} \; \tilde{g}_l^* \tilde{g}_l \right). \tag{4.33}$$

According to the error model in Section 4.3.1.2, the complex gain errors are circular symmetric complex Gaussian distributed with variance $\sigma_g^2$. Hence,

$$\frac{\tilde{g}_l}{\sigma_g} \; \sim \; \mathcal{CN}(0, 1), \quad l = 1, \ldots, L, \tag{4.34}$$

and

$$\frac{2}{\sigma_g^2} \sum_{l=1}^{L} \tilde{g}_l^* \tilde{g}_l \;\; \sim \;\; \mathcal{X}_{2L}^2. \tag{4.35}$$

In analogy to the method in [5], a confidence level, $\alpha_g$, and a corresponding upper confidence bound $\tilde{\mathcal{X}}_{u,\alpha_g}^2$, can be defined. Then, an upper bound can be obtained by $\tilde{\mathcal{X}}_{u,\alpha_g}^2 \hat{\sigma}_g^2$, where $\sigma_g = p_g \, g_0/2$ is chosen according to Section 4.3.1.2. Hence,

$$\mathbb{E}_{\mathbf{Q}} \; \|\mathbf{Q}\|_f^2 \;\; \leq \;\; TK \, \hat{\sigma}_u^2 \, \tilde{\mathcal{X}}_{u,\alpha_g}^2 \frac{\sigma_g^2}{2} \;\; \triangleq \;\; \hat{q}_{\mathrm{MSE}} \; . \tag{4.36}$$

For the remaining term, $\mathbb{E}_{\mathbf{Q}} \; \|\mathbf{Q}\|_f$ in (4.30), an upper bound can be obtained by applying 'Jensen's inequality' [126], i.e.

$$\mathbb{E}_{\mathbf{Q}} \; \sqrt{\|\mathbf{Q}\|_f^2} \;\; \leq \;\; \sqrt{\mathbb{E}_{\mathbf{Q}} \; \|\mathbf{Q}\|_f^2} \;\; \leq \;\; \sqrt{\hat{q}_{\mathrm{MSE}}} \;\; \triangleq \;\; \hat{q}_{\mathrm{RMSE}} \, , \tag{4.37}$$

since the square root is a concave function.

Finally, using (4.14),(4.16) and (4.36)-(4.37), an expression for the right hand side in (4.30) is obtained, yielding

$$\hat{\beta}_{\text{reg}} \quad = \quad \hat{q}_{\text{MSE}} + \hat{n}_{\text{MSE}} + 2\,\hat{n}_{\text{RMSE}}\,\hat{q}_{\text{RMSE}}\,. \tag{4.38}$$

This result is used for regularization.

### 4.5.1 Simulations

In subsequent simulations, 'SPARSE' is compared to *Robust Capon Beamforming* (R-Capon) [107, 108] and *Weighted MUSIC* (W-MUSIC) [109]. The results of a standard *delay-and-sum beamformer* (DSB) [64] are shown as a reference for a non-robust method. The regularization parameter, $\beta_{\text{reg}}$, is estimated according to Section 4.5.

#### 4.5.1.1 Setup

The simulation setup is almost the same as in Section 4.4.1.1.
However, a uniform linear array with $L = 30$ sensors is considered and the sources are located at angles $\vartheta_1 = -45°$ and $\vartheta_2 = -42°$, which corresponds to a spacing of 3° and falls inside the Rayleigh resolution limit in (4.25). The number of sources, $K$, and the source power, $\sigma_u^2$, are assumed to be known. Source correlation is indicated in terms of the correlation coefficient, $\zeta_c$, with $0 \leq \zeta_c \leq 1$ being defined as in (4.26). The noise power, $\sigma_n^2$, is estimated while no source is present. The nominal sensor gain is set to $g_0 = 1$ and the standard deviation of the gain errors is given by $\sigma_g = p_e \frac{g_0}{2}$, where different values of $p_e$, $0 \leq p_e \leq 1$, are considered. Using typical values in [128, 129], the relative gain variance, $\sigma_g^2/g_0^2$, is in the range $[-42, -11]$ dB, which corresponds to $p_e \in [0.02, 0.56]$. The confidence levels are equally chosen to $\alpha_\chi = \tilde{\alpha}_\chi = \alpha_g = 0.9$.

#### 4.5.1.2 DOA Spectra

In Figure 4.4, the DOA spectra obtained by the different methods are depicted. The peaks of the spectral magnitude represent the estimated DOAs. Figures 4.4(a)-4.4(f) show some scenarios for SNR = 15 dB. Figures 4.4(a)-4.4(c) depict the cases without correlation ($\zeta_c = 0$), where W-MUSIC and SPARSE are able to correctly estimate the DOAs. R-Capon is sometimes unstable and yields incorrect peak amplitudes.

Figure 4.4(c) shows an example for improper regularization, taking only noise but no model errors into account. As a result, SPARSE shows several spurious peaks, which misleadingly indicates additional sources. In Figures 4.4(d)-4.4(f), the two source signals are correlated. W-MUSIC and R-Capon suffer from correlation even without model errors, while SPARSE can still resolve both sources for $\zeta_c = 1$ if $p_e = 0$. However, the joint impacts of correlation and model errors also lower the performance of SPARSE. When $\zeta_c = 0.3$ and $p_e = 0.3$, SPARSE, W-MUSIC and R-Capon perform fairly well. For larger values of $\zeta_c$ and $p_e$, the two peaks in the spectrum of SPARSE start to merge and only one source appears (Figure 4.4(f)). Figures 4.4(g)-4.4(i) show results for SNR = 0 dB. At this level, only SPARSE can resolve the two sources for $p_e = 0.65$. However, the accuracy is reduced and the peaks start to merge when $p_e$ further increases.

#### 4.5.1.3 Performance Evaluation

The performance is evaluated in Figure 4.5, using 500 Monte Carlo trials. The upper part of each sub-figure shows the approximate *root mean-squared angular error* (RMSE) in degrees, relative to the grid accuracy, $\delta\vartheta = 1°$. The error is calculated based on the positions (indices) of the significant elements in $\mathbf{x}$, i.e the elements with the largest modulus contained in $\hat{\mathcal{S}}$. In particular, the estimates in $\hat{\mathcal{S}}$ with the smallest distance to the true source locations in $\mathcal{S}$ are used to calculate the error. The lower parts of each figure show the success rates. In a successful trial, the regularization is appropriately chosen. That is, the spectrum exhibits exactly two peaks, corresponding to the two simulated sources (the number of sources is assumed to be knwon for this evaluation). Generally, SPARSE performs better than R-Capon and W-MUSIC. For uncorrelated signals, i.e. $\zeta_c = 0$ (Figures 4.5(a)-4.5(c)), SPARSE works fairly reliable up to gain/phase errors of $p_e = 0.65$. The impact of correlation is shown in Figures 4.5(d)-4.5(f). While W-MUSIC and R-Capon become unstable, SPARSE is merely affected by correlation alone but its robustness against gain/phase errors is reduced. However, for the successful trials, the angular error is still small. At low success rates, less Monte Carlo trials are available, since the error is only calculated in case of a successfull trial. Therefore stronger fluctuations in the RMSE can be observed.

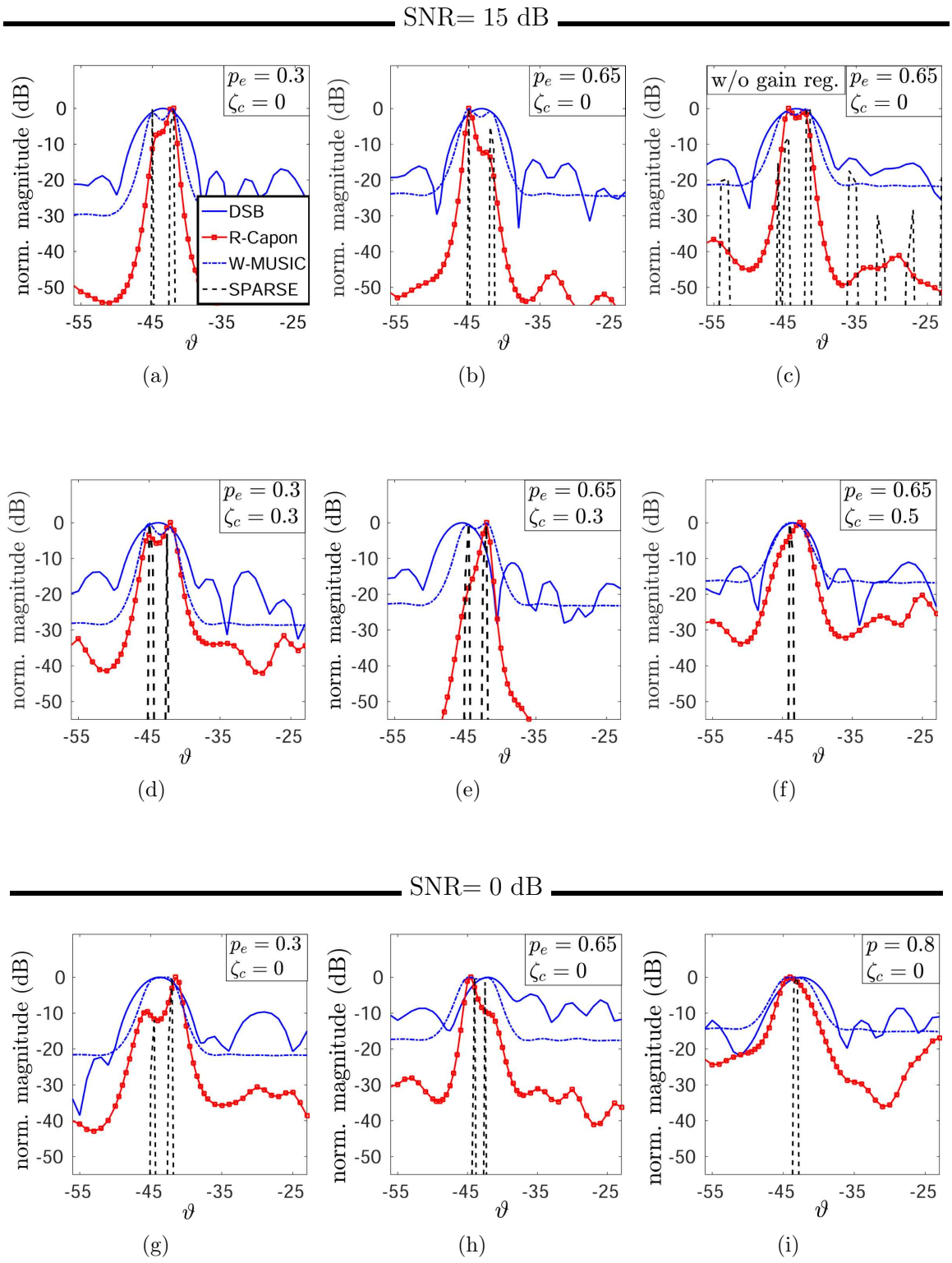



Figure 4.4. Comparison of the DOA spectra obtained by SPARSE, W-MUSIC and R-Capon. The results of a DSB are shown as a reference. For SPARSE, $\hat{\beta}$ was estimated according to Section 4.5.

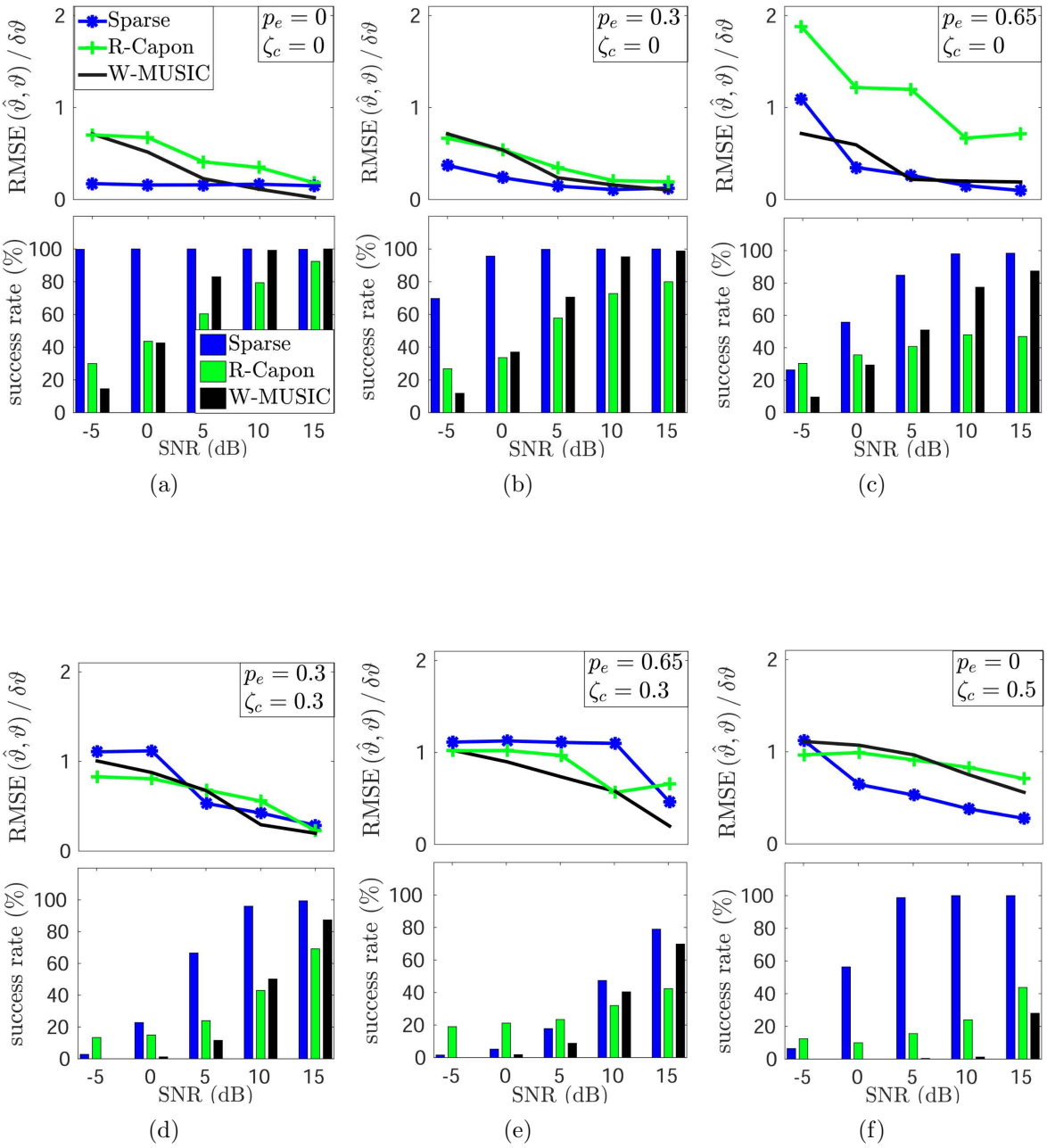



Figure 4.5. Performance of SPARSE, W-MUSIC and R-Capon: success rates (lower subfigures) and angular errors relative to the grid accuracy, $\delta\vartheta$, (upper subfigures) vs. different SNRs. For SPARSE, $\hat{\beta}$ was estimated according to Section 4.5.

## 4.6 Dictionary Calibration for Direction Finding

The material presented in this section is partly taken from [68][9].

Dictionary calibration is performed by applying the robust steering vector estimation methods in [124] to the individual dictionary atoms. Based on the general model in (4.2), the aim is to estimate the perturbed dictionary, $\tilde{\mathbf{D}} = \mathbf{D} + \Delta\mathbf{D}$, which includes any model errors. Each column (atom) of $\mathbf{D}$ contains the ideal steering vector for a certain angle, i.e. $\mathbf{d}_i = \mathbf{d}(\vartheta_i)$, $i = 1, \ldots, N$. For some element $\vartheta \in \{\vartheta_1, \ldots, \vartheta_N\}$, let the error term in the corresponding dictionary atom be denoted by $\Delta\mathbf{d}$. The perturbed dictionary atoms (steering vectors) are obtained by estimating $\Delta\mathbf{d}$ for each angle of interest, i.e.

$$\tilde{\mathbf{d}}(\vartheta) = \mathbf{d}(\vartheta) + \Delta\mathbf{d}\,. \tag{4.39}$$

According to [108], $\Delta\mathbf{d}$ is confined within an ellipsoidal uncertainty set, such that

$$(\tilde{\mathbf{d}}(\vartheta) - \mathbf{d}(\vartheta))^{\mathrm{H}} \mathbf{C}_{\Delta}^{-1} (\tilde{\mathbf{d}}(\vartheta) - \mathbf{d}(\vartheta)) \;\leq\; 1\,. \tag{4.40}$$

The coefficients of the ellipsoid are contained in the matrix $\mathbf{C}_{\Delta}$. A linear transformation can be used in (4.40) to obtain $\mathbf{C}_{\Delta} = \epsilon_{\Delta}\mathbf{I}$ [108], such that (4.40) becomes

$$\|\tilde{\mathbf{d}}(\vartheta) - \mathbf{d}(\vartheta)\|_2^2 \;=\; \|\Delta\mathbf{d}\|_2^2 \;\leq\; \epsilon_{\Delta}\,. \tag{4.41}$$

*A priori* knowledge of $\epsilon_{\Delta}$ is often not available. If the tolerance level is set too high, closely-spaced sources cannot be resolved. The calibration method in [124] is able to estimate the the perturbed steering vectors without requiring a specific value for $\epsilon_{\Delta}$. This method can be adopted for dictionary calibration to estimate $\Delta\mathbf{d}$. Therefore, the method in [124] and some concepts in [108] are briefly reviewed below.
At first, the covariance matrix of the sensor snapshots is modeled by [108]

$$\mathbf{R} = \sum_{k=1}^{K} \sigma_{u,k}^2 \mathbf{d}(\vartheta_k)\mathbf{d}^{\mathrm{H}}(\vartheta_k) + \mathbf{Z}_n\;, \tag{4.42}$$

where $\sigma_{u,k}^2$ denotes the source power of the $k$-th source, $k = 1, \ldots, K$, and $\mathbf{Z}_n$ denotes the noise covariance matrix. Using singular value decomposition on $\mathbf{R}$, one obtains [108]

$$\mathbf{R} \;=\; \mathbf{U}_{\mathbf{R}} \left( \mathbf{\Sigma}_{\mathbf{R}} [\mathbf{I}_K, \mathbf{0}]^{\top} + \mathbf{\Sigma}_{\mathbf{R}} [\mathbf{0}, \mathbf{I}_{L\text{-}K}]^{\top} \right) \mathbf{V}_{\mathbf{R}}^{\mathrm{H}}. \tag{4.43}$$

Herein, $\mathbf{I}_J$ is the $J$-dimensional identity matrix, $\mathbf{U}_{\mathbf{R}}$ and $\mathbf{V}_{\mathbf{R}}$ are unitary matrices, and $\mathbf{\Sigma}_{\mathbf{R}}$ is the diagonal matrix of singular values, where the first $K$ entries belong to the

[9] C. Weiss and A. M. Zoubir, "Robust High-Resolution DOA Estimation with Array Pre-Calibration," in *Proc. of the 22nd European Signal Processing Conference (EUSIPCO)*, Lisbon, Portugal, September, 2014.

source signals and the remaining $L-K$ entries belong to the noise components. Then, the signal subspace can be written by $\mathbf{U}_s = \mathbf{U}_{\mathbf{R}}\mathbf{\Sigma}_{\mathbf{R}}[\mathbf{I}_K, \mathbf{0}]^\top$, and the noise subspace becomes $\mathbf{U}_n = \mathbf{U}_{\mathbf{R}}\mathbf{\Sigma}_{\mathbf{R}}[\mathbf{0}, \mathbf{I}_{L\text{-}K}]^\top$. Further, let $\widehat{\mathbf{U}}_n$ be an estimate of $\mathbf{U}_n$, that is obtained using the sample covariance, $\widehat{\mathbf{R}} = \frac{1}{T}\mathbf{Y}\mathbf{Y}^{\mathrm{H}}$, and knowledge of the dimensionality of the noise subspace, *L–K*, where $K$ number of sources and $L$ is the number of sensors. Then, following the notation in [124], the true noise subspace is given by $\mathbf{U}_n = \widehat{\mathbf{U}}_n + \delta\mathbf{U}_n$, where $\delta\mathbf{U}_n$ is a stochastic estimation error. As it is generally assumed in subspace methods, the steering vectors corresponding to the source signals are orthogonal to the noise subspace [124], i.e.

$$\mathbf{U}_n\tilde{\mathbf{d}}(\vartheta) \;=\; (\widehat{\mathbf{U}}_n + \delta\mathbf{U}_n)^{\mathrm{H}}(\mathbf{d}(\vartheta) + \Delta\mathbf{d}) = \mathbf{0}\,. \tag{4.44}$$

When only an estimate, $\widehat{\mathbf{U}}_n$, is available, the MSE with respect to the stochastic estimation error, $\delta\mathbf{U}_n$, can be approximated by [124]

$$\begin{aligned}\mathbb{E}_{\delta\mathbf{U}_n}\,\|\widehat{\mathbf{U}}_n^{\mathrm{H}}(\mathbf{d}(\vartheta)+\Delta\mathbf{d})\|_2^2 \;&=\; (\mathbf{d}(\vartheta)+\Delta\mathbf{d})^{\mathrm{H}}\left(\mathbb{E}_{\delta\mathbf{U}_n}\,\delta\mathbf{U}_n\delta\mathbf{U}_n^{\mathrm{H}}\right)(\mathbf{d}(\vartheta)+\Delta\mathbf{d})\\ &\approx\; \mathbf{d}(\vartheta)^{\mathrm{H}}\,\mathbf{Z}_\delta\,\mathbf{d}(\vartheta) \;\triangleq\; \beta_\delta^2\,.\end{aligned} \tag{4.45}$$

where $\mathbf{Z}_\delta = \mathbb{E}_{\delta\mathbf{U}_n}\,\delta\mathbf{U}_n\delta\mathbf{U}_n^{\mathrm{H}}$. The approximation in (4.45) is due to the assumption that all products involving $\mathbf{Z}_\delta$ and $\Delta\mathbf{d}$ are small, such that any terms involving their product can be neglected and $\beta_\delta$ depends only on the covariance of the estimation error, $\delta\mathbf{U}_n$. Then, the calibrated dictionary atoms, $\tilde{\mathbf{d}}(\vartheta) = (\mathbf{d}(\vartheta) + \Delta\mathbf{d})$ can be estimated in terms of a quadratic program with quadratic constraints [124],

$$\min_{\Delta\mathbf{d}} \|\Delta\mathbf{d}\|_2^2 \qquad \text{s.t.} \qquad \|\widehat{\mathbf{U}}_n^{\mathrm{H}}(\mathbf{d}(\vartheta)+\Delta\mathbf{d})\|_2^2 \;\leq\; \beta_\delta^2. \tag{4.46}$$

Using Lagrange multipliers, a closed-form solution can be obtained [124]:

$$\widehat{\Delta\mathbf{d}} \;=\; \left(\beta_\delta\left(\mathbf{d}(\vartheta)^{\mathrm{H}}\widehat{\mathbf{U}}_n^{\mathrm{H}}\widehat{\mathbf{U}}_n\mathbf{d}(\vartheta)\right)^{-\frac{1}{2}} - 1\right)\widehat{\mathbf{U}}_n\widehat{\mathbf{U}}_n^{\mathrm{H}}\mathbf{d}(\vartheta). \tag{4.47}$$

When a good estimate of the covariance matrix is available, e.g. when $T$ is large, then $\delta\mathbf{U}_n$ shrinks close to zero and also $\beta_\delta$ tends to zero. It is stated in [124] that, under these conditions, the final result for estimating the steering vectors equals the one obtained by the projection approach in [123]. Using this result, the error in the dictionary atoms, $\Delta\mathbf{d}$, and, hence, the true dictionary atoms, $\tilde{\mathbf{d}}_i(\vartheta_i)$, can be estimated for all considered angles $\vartheta \in \{\vartheta_1, \ldots, \vartheta_N\}$. Hence, the fully calibrated dictionary can be calculated by

$$\widehat{\mathbf{D}} = \left(\mathbf{I} - \widehat{\mathbf{U}}_n\widehat{\mathbf{U}}_n^{\mathrm{H}}\right)\mathbf{D}, \tag{4.48}$$

where $\mathbf{D}$ is the ideal dictionary without phase or gain errors. When $K$ is known, the calibrated dictionary, $\widehat{\mathbf{D}}$, is only based on the subspace estimates. After dictionary calibration, the regularization parameter for sparse estimation can be chosen to account for noise only.

### 4.6.1 Simulations

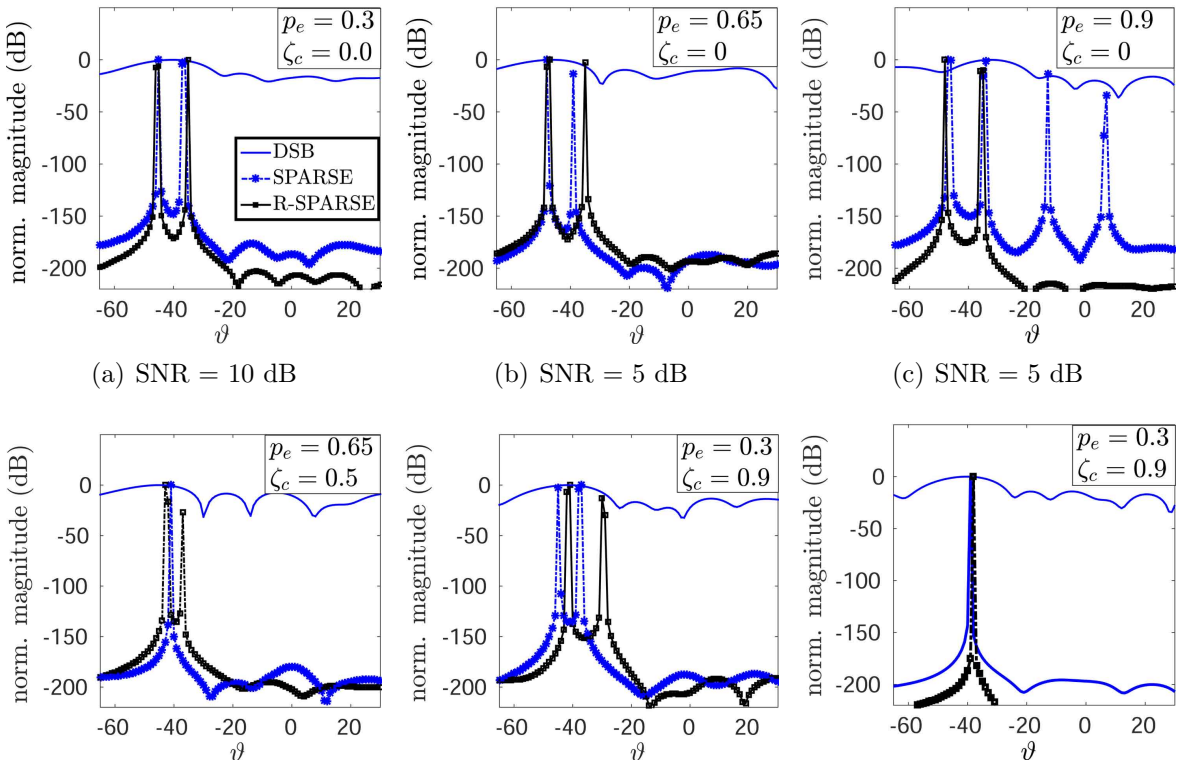
In subsequent simulations, the SPARSE method without dictionary calibration is compared to the case where the calibrated dictionary is used. The latter method is referred to as 'R-SPARSE'. In both cases, the regularization parameter takes only the measurement noise into account.

#### 4.6.1.1 Setup

The simulation setup resembles that of Section 4.4.1.1. However, SPARSE and R-SPARSE use $T = 10$ snapshots for DOA estimation and $T = 200$ snapshots to estimate the covariance matrix, $\mathbf{R}$, which is used for dictionary calibration. Since the calibration procedure within R-SPARSE relies on an accurate estimate of $\mathbf{R}$, an insufficient number of noise samples for its estimation would reduce the performance and robustness. For all simulations, the gain and phase errors are equally set to $p_e = p_\phi = p_g$. In order to determine the regularization parameter (regarding measurement noise only), the confidence levels are equally set to $\alpha_S = \alpha_\mathcal{X} = \tilde{\alpha}_\mathcal{X} = 0.996$ for SPARSE, and $\alpha_R = \alpha_\mathcal{X} = \tilde{\alpha}_\mathcal{X} = 0.5$ for R-SPARSE. Correlation between the complex amplitudes of the source signals is indicated in terms of the correlation coefficient, $\zeta_c$, where $0 \leq \zeta_c \leq 1$. It is defined according to (4.26).

#### 4.6.1.2 DOA Spectra

Figure 4.6 illustrates the DOA spectra obtained by both methods. Besides SPARSE and R-SPARSE, the DSB spectra are shown as a reference. Generally, high robustness against phase and gain errors can be observed for R-SPARSE in all scenarios with and without correlation and even at low SNRs. Figures 4.6(a)-4.6(c) show the uncorrelated case, i.e. $\zeta_c = 0$. Both SPARSE and R-SPARSE can resolve the two sources but severe phase and gain errors with $p_e = 0.9$ can no longer be handled by SPARSE without calibration. The spurious peaks in the spectrum in Figure 4.6(c) misleadingly indicate additional sources. Also, SPARSE is more sensitive to the joint effects of correlation and gain/phase errors than R-SPARSE. In Figure 4.6(d), it is shown that R-SPARSE can deal with moderate correlation of $\zeta = 0.5$ and phase/gain mismatches of $p_e = 0.65$, even at SNR= 5 dB. However, stronger correlation also degrades the performance of R-SPARSE, especially at low SNRs, which is depicted in Figures 4.6(e)-4.6(f).




(a) SNR = 10 dB (b) SNR = 5 dB (c) SNR = 5 dB

(d) SNR = 5 dB (e) SNR = 15 dB (f) SNR = 5 dB

Figure 4.6. Comparison of the DOA spectra obtained by SPARSE and R-SPARSE for different values of $p_e$ and $\zeta_c$. The results of a DSB are shown as a reference. For R-SPARSE, dictionary calibration was performed according to Section 4.6.

#### 4.6.1.3 Performance Evaluation

Figure 4.7 shows the performance of SPARSE and R-SPARSE in various scenarios of different SNRs, correlation levels and gain/phase mismatches, averaged over 500 Monte Carlo trials. The upper part of each figure depicts the angular error, relative to the grid accuracy of $\delta\vartheta = 1°$. The error is calculated based on the positions (indices) of the significant elements in $\mathbf{x}$, i.e the elements with the largest modulus contained in $\hat{\mathcal{S}}$. In particular, the estimates in $\hat{\mathcal{S}}$ with the smallest distance to the true source locations in $\mathcal{S}$ are used to calculate the error. When the success rate is low, less Monte Carlo trials are available. Therefore stronger fluctuations in the RMSE can be observed. The lower parts of each figure show the success rates. In a successful trial, the regularization is appropriately chosen. That is, the spectrum exhibits exactly two peaks, corresponding to the two simulated sources (the number of sources is assumed to be knwon for this evaluation).

For $\zeta_c = 0$ (Figures 4.7(a)-4.7(c)), the performance of SPARSE is limited to a narrow range of SNRs when the confidence level, $\alpha_S$, is fixed. At lower SNRs, one peak is

often suppressed and at high SNRs, the regularization is too weak, yielding various spurious peaks in the spectrum. R-SPARSE, in contrast, yields higher success rates and lower angular errors at all SNRs and for all considered values of $p_e$. When the source signals are correlated (Figures 4.7(d)-4.7(f)), the joint impact of gain/phase errors and correlation reduce the performance of both methods. This effect is more notable for SPARSE but R-SPARSE requires higher SNRs to achieve low angular errors in the presence of correlation. The success rates are only affected at lower SNRs. For SPARSE, the angular error increases significantly below a certain SNR threshold and also the success rates are reduced.
In Figures 4.7(g)-4.7(i), a change in the confidence levels $\alpha_S$, $\alpha_R$ is introduced in order to show the robustness of R-SPARSE to the choice of the regularization parameter over a wide range of SNRs. The new values are set to $\tilde{\alpha}_S = \alpha_S + \Delta\alpha_S = 0.91$ and $\tilde{\alpha}_R = \alpha_R + \Delta\alpha_R = 0.8$. They correspond to a small change of $\Delta\alpha_S = 0.086$ for SPARSE, and a larger change of $\Delta\alpha_R = 0.3$ for R-SPARSE. The difference is more emphasized by considering the ratio $\Delta\alpha_R/\Delta\alpha_S$, which is around 11 dB. For a small change in $\alpha_S$, SPARSE achieves good performance only within a narrow SNR region. This region is significantly shifted as compared to that obtained by the original value of $\alpha_S$. For R-SPARSE, in contrast, the confidence level is very different from the original value but the performance is only slightly affected at low SNRs. At higher SNRs, R-SPARSE works reliably in the presence of gain and phase errors at moderate correlation levels.

## 4.7 Computational Complexity

For regularization, the complexity of calculating $\hat{\beta}_{\text{reg}}$ depends on how $\sigma_u^2$ and $\sigma_n^2$ are estimated. Note that in many practical applications also $K$ has to be estimated. For the purpose of this analysis, however, it is assumed to be known. Regarding the problem of dictionary calibration, the computational complexity is determined by the subspace decomposition of $\widehat{\mathbf{R}}$, which is $O(L^3)$ [5].
In both cases, the overall complexity is dominated by the sparse estimation method, i.e. $\ell_1$-minimization. Herein, the complexity increases linearly in $T$ [5, 130]. However, a more efficient algorithm is proposed in [5], where only $O((KN)^3)$ operations and at least $O((KN)^{0.5})$ iterations re required.
For W-MUSIC, the subspace decomposition of the correlation matrix is the dominating factor with complexity $O(L^3)$ [5]. For R-Capon, an optimization problem of complexity $O(L^3)$ needs to be solved [108]. This has to be done for each dictionary atom, which results in a complexity of $O(NL^3)$. Nevertheless, the individual dictionary atoms can be determined separately using parallel processing.

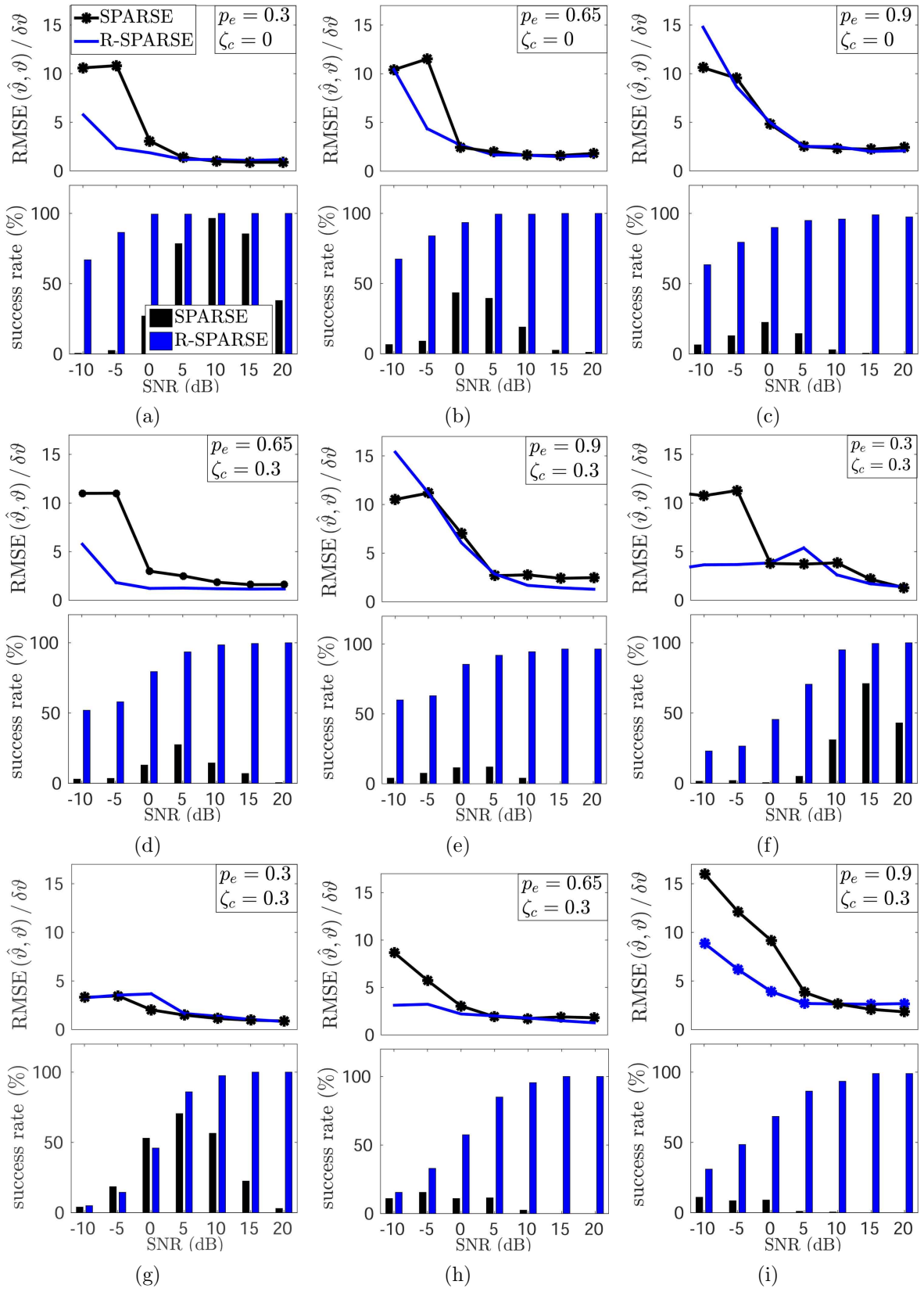



Figure 4.7. Performance of SPARSE and R-SPARSE: success rates (lower subfigures) and angular errors relative to the grid accuracy, $\delta\vartheta$, (upper subfigures) vs. different SNRs. For R-SPARSE, dictionary calibration was performed according to Section 4.6.

## 4.8 Discussion and Findings

The presented regularization methods provide a statistical framework to select the regularization parameter, $\hat{\beta}_{\text{reg}}$, for $\ell_1$-minimization. This can reduce the computational costs compared to a blind search. The joint impacts of model errors and noise are reflected in an inaccurate generative model for creating the sparsity-promoting dictionary. Simulations show that proper regularization ensures stable operation of the SPARSE algorithm under sensor position errors and general gain/phase mismatches. The obtained estimates of $\beta_{\text{reg}}$ represent a suitable tolerance level to avoid spurious peaks in the spectrum without suppressing the true sources. This kind of regularization can also be used in different applications such as radar, sonar or communications, where the problem of direction finding arises.
The parameters $\alpha_\chi, \tilde{\alpha}_\chi, \alpha_\phi$ and $\alpha_g$ can be set in advance. Based on these choices, the upper and lower confidence bounds $\mathcal{X}_{u,\alpha_\chi}, \mathcal{X}_{l,\alpha_\chi}$, $\mathcal{X}_{u,\tilde{\alpha}_\chi}, \mathcal{X}_{u,\tilde{\alpha}_\chi}$ as well as $\mathcal{N}_{u,\alpha_\phi}$ and $\mathcal{X}^2_{u,\alpha_g}$ can be determined prior to system operation and stored in a database. The parameters $\sigma_\phi$ and $\sigma_g$ can be set by the user, based on available information of the system. For example, this information can be based on the specifications of the sensors and the surrounding circuitry, or it can be acquired from system monitoring (e.g. temperature). The variables $\sigma_\phi$ and $\sigma_g$ are directly related to the system parameters, such that regularization can be adapted to account for the specific impairments at hand. It is also possible to incorporate other system-specific error terms in the presented framework under the premise that their mutual independence with other model errors can be assumed.
For W-MUSIC and R-Capon several snapshots are required to estimate the covariance matrix, while SPARSE can operate with a single snapshot. Nevertheless, additional snapshots improve the SNR conditions and, hence, the performance. Moreover, in contrast to W-MUSIC and R-Capon, SPARSE does not make assumptions about the correlation of the source signals. Therefore, the methods cannot be directly compared when the source signals are correlated. Generally, SPARSE is more robust and accurate than W-MUSIC and R-Capon in the presence of model errors, which comes at the cost of a higher computational complexity. However, simulation show that the joint impacts of model errors and correlation also degrade the performance of SPARSE at low SNRs. Therefore, correlation impacts its robustness to model errors.
Drawback of the regularization methods is the need for estimating the source power, $\sigma_s^2$, the noise power, $\sigma_n^2$, and the number of sources, $K$. Although robust methods for the individual terms exist, they can still be inaccurate if all parameters have to be jointly estimated. Also, regularization can only adapt the tolerance level to certain impairments but the accuracy decreases when the model errors are too severe. This is because the quality of the dictionary suffers from an inaccurate generating model.

As a remedy, R-SPARSE uses dictionary calibration in a addition to regularization. Simulations show that reliable results with high success rates over a wide range of SNRs can be obtained using a properly calibrated dictionary. Also, the sparse solution obtained by $\ell_1$-minimization is more robust to the choice of the regularization parameter and even large variations have only minor impact over a wide range of SNRs. This can be ascribed to the fact that the calibration procedure mitigates most of the errors introduced by noise and array imperfections, which is also the reason why R-SPARSE is fairly robust when the source signals are correlated.
However, dictionary calibration requires a good estimate of the covariance matrix. Another drawback is that R-SPARSE requires knowledge of the number of sources. Similar to W-MUSIC and R-Capon, its performance strongly depends on the quality of this estimate.
In an extension of this work, a systematic approach to determine the choice of the confidence levels, $\alpha_g, \alpha_{\mathcal{X}}, \alpha_{\mathcal{X}}$, and $\alpha_\phi$, can be developed. Also, off-grid sources can be taken into account by an adaptive grid-refinement as proposed in [5].

## 4.9 Conclusion

A sparse regularization method for direction finding based on sparse estimation (SPARSE) is presented. It accounts for sensor position errors and general phase/gain mismatches in the generating model of the dictionary. Also, other model errors can be incorporated in this framework if their mutual independence can be assumed. Since the dictionary is derived from an ideal model, the regularization parameter of the sparse estimation problem has to impose a suitable tolerance level to handle the joint impacts of model errors and noise. This parameter is estimated as an upper bound of the MSE between the perturbed model, including errors such as phase and gain mismatches, and the ideal model. Moreover, it is directly connected to certain model errors, such that available prior knowledge of the system can be used to achieve selective regularization of specific errors. Hence, the regularization parameter can be adapted to the current situation in an automated fashion. Also, major parts of its derivation can be done offline and stored in a database.
The SPARSE method with proper regularization shows better robustness and higher resolution in the presence of sensor position errors and noise than the compared robust methods. However, its robustness to model errors is reduced when the source signals are correlated, especially at low SNRs. A drawback of this regularization method is the need for estimating the source power, $\sigma_s^2$, the noise power, $\sigma_n^2$, and the number of sources, $K$. Also, the regularization parameter depends on the confidence levels, $\alpha_g, \alpha_\phi, \alpha_{\mathcal{X}}$, and $\tilde{\alpha}_{\mathcal{X}}$, which are empirically determined. As a remedy, a combined

dictionary calibration and sparse estimation method, R-SPARSE, is presented. The calibrated dictionary takes the joint impacts of model errors and noise into account and provides stable and reliable sparse estimation results. Regularization alone can only adjust the tolerance level but the accuracy is limited by an erroneous dictionary. Therefore, dictionary calibration also improves the accuracy. Although regularization is still required for $\ell_1$-minimization, the problem of choosing the regularization parameter (i.e. the confidence levels) is significantly alleviated. In particular, the sparse solution obtained by R-SPARSE is only marginally affected even by large variations of the confidence level and high success rates achieved for a wide range of SNRs. However, the calibration performance depends on how good the covariance matrix and the number of sources can be estimated. Therefore, dictionary calibration is applicable when many snapshots and knowledge of the number of sources are available. It turns out that the joint impacts of correlation and model errors degrade the performance of both SPARSE and R-SPARSE. However, since R-SPARSE mitigates some model errors by calibration prior to sparse estimation, it is more robust to these effects.
In a future research, a systematic approach to choosing to confidence levels for regularization can be considered. In addition, other types of model errors and off-grid sources can be taken into account. Also, uncertain dictionary parameters such a the speed of propagation or the signal frequency can be estimated. Related dictionary learning techniques are presented in the ensuing chapters.

# Chapter 5

# Alternating Sparse Estimation and Dictionary Learning

This chapter introduces a sparse estimation and dictionary learning framework based on *alternating estimation* (AE). While a sparse representation in the previous chapters is obtained based on a variant of $\ell_1$-minimization, it is subsequently assumed that the employed dictionary is highly redundant, such that the necessary RIP conditions are no longer fulfilled. Therefore, a dictionary pre-processing routine is applied to yield a modified dictionary with improved RIP conditions. Using the modified dictionary, a sparse representation of the signal is estimated using a greedy OMP-based algorithm. In Chapter 4, uncertainty in the dictionary is represented by a general error term, $\Delta\mathbf{D}$. The true dictionary is given by $\mathbf{D} + \Delta\mathbf{D}$, where $\mathbf{D}$ is the dictionary obtained under ideal conditions. Different from this approach, this chapter considers uncertainty in terms of imperfectly known global and local dictionary parameters, $\boldsymbol{\theta}$, that represent particular global and local characteristics of the system. Hence, the dictionary can be written as a function of these parameters, i.e. $\mathbf{D}(\boldsymbol{\theta})$. It is found in Chapter 4, that simply adjusting the tolerance level to account for uncertainty by regularization can improve robustness to model errors but it cannot improve the estimation performance. Therefore, in this chapter, AE-based estimation is used to iteratively estimate the dictionary parameters along with a sparse representation of the signal. While dictionary calibration in Chapter 4 is a static operation that is performed only once, the pre-processing routine in this chapter is applied in each AE iteration and the estimate of $\boldsymbol{\theta}$ is successively improved. The generic sparse estimation and dictionary learning framework is specified for the application of fiber-optic sensing in combination with *compressed sensing* (CS) according to Chapter 3.2 and Chapter 3.3, respectively.

The material presented in this chapter is partly taken from [87][1], [88][2], [30][3] [131][4].

[1] C. Weiss and A. M. Zoubir, "Fiber Sensing Using UFWT-Lasers and Sparse Acquisition," in *Proc. of the 21st European Signal Processing Conference (EUSIPCO)*, September, 2013.

[2] C. Weiss and A. M. Zoubir, "Fiber Sensing Using Wavelength-Swept Lasers: A Compressed Sampling Approach," in *Proc. of the 3rd International Workshop on Compressed Sensing Theory and its Applications to Radar, Sonar and Remote Sensing (CoSeRa)*, June, 2015.

[3] C. Weiss and A. M. Zoubir, "A Compressed Sampling and Dictionary Learning Framework for Wavelength-Division-Multiplexing-Based Distributed Fiber Sensing," accepted for publication in *Journal of the Optical Society of America A*, 2017 (assigned issue: vol. 34, no. 5).

[4] C. Weiss and A. M. Zoubir, "Dictionary Learning Strategies for Compressed Fiber Sensing Using a Probabilistic Sparse Model" submitted to *IEEE Transactions on Signal Processing*, 2016.

**Chapter Outline**

Section 5.1 gives an introduction and motivation and Section 5.2 provides an overview of state-of-the-art techniques and related work. In Section 5.3, the problem statement is introduced. Section 5.4 provides theoretical limits for sparse estimation and dictionary learning in terms of the Cramér-Rao bound. In Section 5.5, a unified framework for compressed fiber sensing and dictionary learning is introduced. Statements about the computational complexity of the presented estimation algorithms are made in Section 5.6. Section 5.7 evaluates the applicability of the presented framework and the performance of the estimation algorithms in various scenarios. A discussion of the results and findings is provided in Section 5.8. Section 5.9 gives a conclusion for this chapter.

## 5.1 Introduction and Motivation

Parametric dictionaries are created from a mathematical model for the signal of interest. It is often encountered in practice that some model parameters are unknown. When the dictionary contains uncertainty in terms of unknown parameters, they have to be estimated along with the desired sparse representation. AE-based estimation [52, 53] is a frequently applied technique to achieve this goal. In an iterative process, it alternates between estimating the sparse representation and the dictionary parameters, thereby successively improving the obtained estimates. In the literature, AE is given different names such as 'block-nonlinear Gauss-Seidel method' or 'block coordinate descent' [54]. The performance and reliability of sparse estimation algorithms is limited by the RIP requirements with respect to the dictionary coherence and the sparsity of the signal. Redundancy in the dictionary helps to obtain highly sparse representations but also aggravates the RIP requirements for estimating a unique sparse representation. Moreover, verifying the RIP for a given dictionary is NP hard [25] and requires knowledge of the expected sparsity level. Therefore, it is desirable to have a sparse estimation and dictionary learning algorithm that can readily handle strong dictionary coherence to estimate the desired sparse representation along with the unknown dictionary parameters.

In order to investigate this problem, the general and abstract signal model for sparse estimation is specified for the application of fiber-optic sensing, according to Chapter 3.2. The aim is to obtain an estimate of the FBG reflection delays in an automated fashion. They are used to infer the quantity and nature of perturbations at the FBGs. The received sensor signal is acquired using CS, according to Chapter 3.3. To this end,

a versatile and unified framework for *compressed fiber sensing and dictionary learning* (CFS-DL) is presented. It incorporates a detailed parametric signal model for a common WDM-based quasi-distributed fiber-optic sensor. The model is compiled using existing physical models for the individual components. In particular, a modular system architecture is assumed, where each subsystem can be described by a *linear time-invariant* (LTI) component. Based on this model, a parametric shift-invariant dictionary is created. It is composed of several sub-dictionaries corresponding to the reflections from the individual FBGs. Hence, the indices of the significant components in the sparse signal, $\mathcal{S}$, can be directly used to estimate the reflection delays. Uncertainty in the dictionary is considered in terms of unknown global and local dictionary parameters. The dictionary itself is highly redundant. Compressed sensing-based signal acquisition further increases the redundancy of the combined sensing matrix and aggravates the RIP requirements (see Appendix A.1). To estimate the sparse representation and the dictionary parameters in the presence of strong dictionary coherence, CFS-DL employs an AE-based algorithm for parametric dictionary learning and OMP-based sparse estimation with inter-atom interference mitigation, which is referred to as PDL-OIAI. It is able to deal with strong dictionary coherence by incorporating an *inter-atom interference* (IAI) mitigation sub-routine to yield a modified sensing dictionary that features a significantly lower coherence level. After IAI mitigation, a simple greedy OMP algorithm with low computational complexity can be used to estimate the desired sparse representation. In order to emphasize the impact of dictionary coherence on the estimation performance, a reference method based on the standard OMP algorithm without IAI mitigation, called PDL-OMP, is introduced. Moreover, the Cramér-Rao bound for jointly estimating a sparse representation and a dictionary parameter is derived. This bound is used to evaluate the estimation performance and efficacy of both algorithms by simulations in various scenarios of different dictionary coherence levels, CS matrices and *signal-to-noise-ratios* (SNRs). The parametric model is adapted to match the experimental setup of a real fiber-sensor. Experimental data of the same sensor is used to assess the practical applicability of the CFS-DL framework.

## 5.2 State of the Art and Related Work

The combination of CS and WDM-based quasi-distributed fiber-optic sensing has been considered recently [30, 87, 88, 131]. Analog and digital CS can be distinguished. While the task in digital CS is to efficiently encode a given signal (or an image) [100], analog CS attempts to achieve compression in the analog domain prior to sampling. Herein, serial and parallel hardware architectures can be used [22]. The 'Xampling' framework [22, 103, 132] describes the sampling process by unions of subspaces. The work

in [103] investigates Xampling based on serial and parallel hardware architectures. The authors in [101] attempt to reduce the number of branches in parallel architectures and state problems related to high filter orders and correlation in serial architectures. In this chapter, a parallel architecture is considered. It allows for a lower average sampling rate and can help to reduce the hardware costs by employing several low-rate *analog-to-digital converters* (ADCs) in parallel, rather than a single costly high-speed ADC. There exist different models to describe the acquisition process using CS. Multi-coset sampling, for example, decribes a periodic non-uniform sampling scheme [98,99]. In this chapter, sparse CS matrices based on the *Database-Friendly* (DF) distribution [106] are of particular interest. They describe a non-uniform sampling process, which can be realized at low average sampling rates. Moreover, their sparse structure makes them storage-efficient.

Various methods for designing a parametric dictionary have been proposed. The authors in [35, 41] introduce coherence constraints to control the design process. In [29,105], conventional coherence measures, such as the mutual coherence or the Babel function (or $\ell_1$-coherence), are considered to obtain an optimal sensing dictionary. The performance of these methods can be improved when the measured data is taken into account during the design process of the sensing dictionary [32,105].

One important class of dictionaries are the composite dictionaries [9, 24, 33]. In [24], a theoretical analysis for pairs of orthonormal bases is provided. The works in [9, 33] consider more general types of dictionaries and use the 'spark' of a matrix in relation to other coherence measures. A later study in [27] investigates redundant dictionaries for CS. The dictionary used in the CFS-DL framework is also composed of different sub-dictionaries, corresponding to the reflections of the individual FBGs. Each sub-dictionary has a translation-invariant structure, as in [36–38]. Similar to the dictionaries in [133,134], the dictionary in CFS-DL is created from a mathematical signal model, which is particularly designed for the application at hand. The parameters of the model used to create the dictionary are adapted to match the experimental setup in [70,86]. Unlike the works in [133,134], the composite structure of the dictionary in CFS-DL allows for uncertainty in terms of global and also local dictionary parameters for each sub-dictionary.

For *dictionary learning* (DL), different strategies have been proposed in the literature [45–50, 56, 135]. A Bayesian framework for joint sparse estimation and DL is adopted in [56]. The work in [50] is one of the first sources that report the concept of AE-based estimation. Provable recovery guarantees and convergence bounds have been obtained [52,53]. Among the most popular AE variants for estimating a non-parametric dictionary are the iterative least-squares DL algorithms [45], the K-SVD algorithm [46], and a recursive least-squares algorithm [47]. Also, robust methods have been proposed, e.g. in [48, 49, 135]. There also exist DL methods for parametric dictionaries. In [136],

information criteria are used for jointly estimating the model order and the parameters. A related work in [137] overcomes the problem of parameter quantization, which results from a finite grid of dictionary atoms by using the measured data to adjust the dictionary parameters. The authors follow the AE paradigm, using an $\ell_p$-RLS algorithm for estimating a sparse representation and least-squares optimization with coherence constraints to estimate the dictionary parameters. The work in [39] applies a polar interpolation between neighboring dictionary atoms to overcome the quantization effect and to limit the dictionary coherence. Another work in [138] uses AE for parametric DL by solving an optimization problem in a distributed fashion over a network of sensors. In this chapter, an AE-based algorithm, called PDL-OIAI, is introduced. It estimates a sparse representation based on a greedy OMP algorithm, while parametric DL is performed by minimizing a local cost function. PDL-OIAI is also equipped with a data-based IAI mitigation sub-routine, according to [32], which enables the algorithm to handle strong dictionary coherence. Instead of explicitly imposing restrictions on the coherence level, the IAI mitigation procedure yields a modified dictionary of reduced dictionary coherence.

## 5.3 Problem Statement

The abstract and general problem of sparse estimation is specified for the application of fiber-optic sensing as described in Chapter 3.2, where a composite shift-invariant dictionary is considered. It is created from $K$ concatenated sub-dictionaries and the atoms of the $k$-th sub-dictionary are associated with the delay of the reflection from the $k$-th FBG, $k = 1, \ldots, K$. Uncertainty in the dictionary is described by unknown dictionary parameters, $\boldsymbol{\theta} \in \Theta$, where $\Theta$ is the considered parameter space. There exist local parameters, $\boldsymbol{\theta}_{\mathrm{LO},k}$, and/or global parameters, $\boldsymbol{\theta}_{\mathrm{G}}$, such that $\boldsymbol{\theta} = [\boldsymbol{\theta}_G, \boldsymbol{\theta}_{\mathrm{LO},1}, \ldots, \boldsymbol{\theta}_{\mathrm{LO},K}]^\top$ and $\boldsymbol{\theta}_k = [\boldsymbol{\theta}_G, \boldsymbol{\theta}_{\mathrm{LO},k}]^\top,\ k = 1, \ldots, K$. Local parameters can be related to the mathematical models for the FBG reflections, $r_k(t),\ k = 1, \ldots, K$, while global parameters represent rather general system parameters such as the receiver bandwidth. Using the fiber sensing model in Chapter 3.2 with additional unknown dictionary parameters, the redundant parametric shift-invariant dictionary is given by

$$\mathbf{D}(\boldsymbol{\theta}) \;=\; [\mathbf{D}_1(\boldsymbol{\theta}_1), \ldots, \mathbf{D}_k(\boldsymbol{\theta}_k), \ldots, \mathbf{D}_K(\boldsymbol{\theta}_K)], \tag{5.1}$$

where $D_k(\boldsymbol{\theta}_k)$, corresponds to the reflection from the $k$-th FBG, $k = 1, \ldots, K$. The signal is acquired using CS, according to Chapter 3.3. To this end, the model for fiber-optic sensing in Chapter 3.2 is refined to include CS-based acquisition. Let $\mathbf{r} \in \mathbb{R}^L$ contain $L$ samples of the received sensor signal, $r(t)$. Then, the CS measurements

become

$$\mathbf{y} \;=\; \mathbf{\Phi r} \;=\; \mathbf{\Phi D}(\boldsymbol{\theta})\mathbf{x} + \mathbf{n} \;=\; \mathbf{B}(\boldsymbol{\theta})\mathbf{x} + \mathbf{n}\;, \tag{5.2}$$

where $\mathbf{\Phi}$ is the CS matrix, $\mathbf{B}(\boldsymbol{\theta}) = \mathbf{\Phi D}(\boldsymbol{\theta}) \in \mathbb{R}^{M\times N}$ is the combined sensing matrix, $\mathbf{r}$ is the observed sensor signal, and $\mathbf{n}$ is a zero mean Gaussian noise term with independent and identically distributed entries, $n_l \sim \mathcal{N}(0, \sigma_n^2)$, $l = 1, \ldots, L$. Hence, the likelihood function is given by

$$p(\mathbf{y}\,|\,\mathbf{x}, \boldsymbol{\theta}) \;=\; \left(2\pi\sigma_n^2\right)^{-\frac{1}{2}} \exp\!\left(-\frac{1}{2\sigma_n^2}\|\mathbf{y} - \mathbf{B}(\boldsymbol{\theta})\mathbf{x}\|_2^2\right). \tag{5.3}$$

Each row of $\mathbf{\Phi}$ describes the measurements taken during one laser sweep. It is assumed that the perturbation profile of the fiber is stationary for $M$ sweeps (until the CS acquisition process is completed). Denoting the laser sweep rate by $S_\mathrm{r}$, the CS acquisition time is given by $M/S_\mathrm{r}$.
The ultimate goal is to estimate the dictionary parameters, $\boldsymbol{\theta}$, along with a sparse representation, $\mathbf{x}$, that correctly indicates the delays of the FBG reflections.

## 5.4 Theoretical Performance Limits: The Cramér-Rao Bound

In this section, the theoretical limits of the achievable estimation performance are discussed. To this end, the *Cramér-Rao bound* (CRB) is derived, which gives a lower bound for the variance of any unbiased estimator and, hence, for the *mean-squared error* (MSE). In the scope of this derivation, it is assumed that the dictionary contains one unknown global parameter, $\theta \in \Theta$, such that $\dim(\Theta) = 1$.

First, the 'constrained CRB' is reviewed according to [139]. It gives a lower bound for estimating the sparse coefficients, $\mathbf{x}$, when $\theta$ is known. Next, the CRB for $\theta$ is derived for a known $\mathbf{x}$. Finally, the mutual information shared between $\mathbf{x}$ and $\theta$ is taken into account and the CRB for jointly estimating the tuple $(\mathbf{x}, \theta)$, is derived.

### 5.4.1 Estimating a Sparse Representation

This part is a review of the basic ideas and concepts used for the derivation in [139], which also includes some necessary prerequisites for subsequent derivations.

The constrained CRB represents a lower bound for estimating the sparse vector $\mathbf{x}$

in (5.2), constrained to a certain class of valid estimators. It is assumed that the dictionary parameter, $\theta$, is known and only $\mathbf{x}$ has to be estimated. The constrained CRB is a local bound that is only valid for estimating vectors in a certain neighborhood of a target point, $\mathbf{x}_0 \in \mathbb{R}^N$. This neighborhood is specified by an $\epsilon$-environment around $\mathbf{x}_0$, i.e. $B_\epsilon(\mathbf{x}_0) = \{\mathbf{x} \in \mathbb{R}^N \,|\, \|\mathbf{x} - \mathbf{x}_0\|_2 < \epsilon\}$, and by a 'locally balanced' constraint set, $\mathcal{X}_c$. According to [139], the constrained set in the sparse setting cannot be written in the form used for the classical constrained CRB. Therefore, the locally balanced constrained set is introduced. The term 'locally balanced' means that, for a certain point $\mathbf{x}_0 \in \mathcal{X}_c$, and another point $\mathbf{x} \in \widetilde{\mathcal{X}}_c \subset \mathcal{X}_c$, this set is locally defined at $\mathbf{x}_0$, such that

$$\mathbf{x}' \;=\; \mathbf{x}_0 + \lambda(\mathbf{x} - \mathbf{x}_0) \;\in\; \widetilde{\mathcal{X}}_c \qquad \forall\, |\lambda| \leq 1\,. \tag{5.4}$$

Herein, the vectors $(\mathbf{x}-\mathbf{x}_0)/\|(\mathbf{x}-\mathbf{x}_0)\|_2,\ \mathbf{x} \in \widetilde{\mathcal{X}}_c$, belong to the set of 'feasible directions' at the point $\mathbf{x}_0$. These are the directions in which one can move without violating the constraints. An estimator for which the constrained CRB is a valid bound must satisfy unbiasedness within the neighborhood $\{\mathbf{x} \in \mathcal{X}_c \,|\, \|\mathbf{x} - \mathbf{x}_0\|_2 < \epsilon\}$, i.e. in the $\epsilon$-environment around $x_0$ where the constraints $\mathcal{X}_c$ are fulfilled. Therefore, the bias gradient has to vanish with respect to the feasible directions, where the constraints are not violated.

The feasible directions are important for the concept of $\mathcal{X}_c$-unbiasedness. It is shown in [139] that $\mathcal{X}_c$-unbiasedness is in fact a property of the subspace spanned by the feasible directions. Any orthonormal basis can be chosen to describe this subspace but its dimensionality may change for different points $\mathbf{x}_0$. It is convenient to use the canonical basis in a finite-dimensional Euclidean space, defined by vectors $\{\mathbf{e}_d\}_{d=1}^D$, which can be collected in a matrix $\mathbf{U}_f = [\mathbf{e}_1, \ldots, \mathbf{e}_D]$. For notational convenience, the dependence of $\mathbf{U}_f$ on $\mathbf{x}_0$ is not explicitly mentioned below. Then, for an estimator, $\hat{\mathbf{x}}$, with associated bias function, $\mathbf{b}(\mathbf{x}) = \mathbb{E}_{\hat{\mathbf{x}}}\,\hat{\mathbf{x}} - \mathbf{x},\ \mathbf{x} \in \mathbb{R}^N$, the term '$\mathcal{X}_c$-unbiasedness' means that the bias gradient vanishes with respect to the subspace spanned by the feasible directions, i.e. $\mathbf{U}_f(\frac{\partial \mathbf{b}(\mathbf{x})}{\partial \mathbf{x}}) = \mathbf{0}$ [139]. This is a requirement for all points $\mathbf{x} \in \{\,\mathcal{X}_c \cap B_\epsilon(\mathbf{x}_0)\,\}$. In other words, the class of valid estimators for which the constrained CRB applies is required to have a vanishing bias gradient for all $\mathbf{x} \in \{\,\mathcal{X}_c \cap B_\epsilon(\mathbf{x}_0)\,\}$. Such estimators are called $\mathcal{X}_c$-unbiased [139]. For a vanishing bias gradient, and for a $K$-sparse target vector, $\mathbf{x}$, the constraint CRB takes the form [139]

$$\mathrm{Cov}(\hat{\mathbf{x}}) \;\succeq\; \mathbf{U}_f \left(\mathbf{U}_f^\top \boldsymbol{\mathcal{I}}(\mathbf{x}) \mathbf{U}_f\right)^{-1} \mathbf{U}_f^\top \;=\; \mathbf{U}_f \boldsymbol{\mathcal{I}}_K^{-1}(\mathbf{x}) \mathbf{U}_f^\top\,, \qquad \|\mathbf{x}\|_0 = K, \tag{5.5}$$

where $\boldsymbol{\mathcal{I}}(\mathbf{x})$ is the *Fisher Information matrix* (FIM), $\boldsymbol{\mathcal{I}}_K(\mathbf{x}) = \mathbf{U}_f^\top \boldsymbol{\mathcal{I}}(\mathbf{x}) \mathbf{U}_f$ is the $K$-reduced FIM, and $\mathbf{U}_f$ contains the directions corresponding to the non-zero entries in $\mathbf{x}_0$ at the indices in $\mathcal{S}$. A detailed derivation of $\boldsymbol{\mathcal{I}}(\mathbf{x})$ can be found in Appendix A.2. According to [139, 140], the FIM and the $K$-reduced FIM are given by

$$\boldsymbol{\mathcal{I}}(\mathbf{x}) = \frac{1}{\sigma_n^2} \mathbf{B}^\top \mathbf{B}, \qquad \text{and} \qquad \boldsymbol{\mathcal{I}}_K(\mathbf{x}) = \frac{1}{\sigma_n^2} \mathbf{B}_\mathcal{S}^\top \mathbf{B}_\mathcal{S}, \tag{5.6}$$

where $\sigma_n^2$ is the noise variance, $\mathbf{B}$ is the combined sensing matrix in (5.2), and $\mathbf{B}_\mathcal{S}$ is a sub-matrix of $\mathbf{B}$ that is composed of the columns with indices in $\mathcal{S}$. The dictionary parameters, $\boldsymbol{\theta}$, are omitted, since they are assumed to be fixed in this context.
In the considered application of FBG-based fiber-optic sensing, it is assumed that the sparsity level, $K$, is known. This is a valid assumption in the absence of noise, since $K$ is the number of reflections, i.e. the number of FBGs. When this assumption is relaxed and the signal is allowed to be $s$-sparse with $0 \leq s \leq K$, then the constrained CRB is the same as in the unconstrained case [139]. This result is related to the fact that inequality constraints do not alter the value of the CRB, as stated in [141].

Some properties of the constrained CRB are stated in [139] and summarized below:

(i) The constrained CRB can be lower than the unconstrained version, where the bias gradient has to vanish for all possible directions. The latter is a stronger requirement, imposing additional restrictions on the class of estimators.

(ii) A requirement for the existence of (5.5) is that the range space $\mathcal{R}(\mathbf{U}_f\mathbf{U}_f^\top)$ has to be a subset of $\mathcal{R}(\mathbf{U}_f\mathbf{U}_f^\top\boldsymbol{\mathcal{I}}(\mathbf{x})\mathbf{U}_f\mathbf{U}_f^\top)$. When $\text{spark}(\mathbf{B}) > 2K$, then $\mathbf{B}$ has unique reconstruction properties for any $K$-sparse vector and $\mathbf{U}_f^\top\boldsymbol{\mathcal{I}}(\mathbf{x})\mathbf{U}_f$ is invertible [9, 139]. Hence, the range spaces are equal and the requirement is fulfilled.

As stated in [139], the best achievable performance of $\mathcal{X}_c$-unbiased estimators is that of the oracle estimator, which has perfect knowledge of the support. Therefore, this estimator would asymptotically achieve the variance proposed by the constrained CRB.

#### 5.4.1.1 The constrained CRB for orthonormal sub-matrices

According to (5.5)-(5.6), the constrained CRB yields a lower bound for the MSE, given by [139]

$$\text{MSE}(\hat{\mathbf{x}}, \mathbf{x}) \;=\; \mathbb{E}\;\|\mathbf{x} - \hat{\mathbf{x}}\|_2^2 \;\geq\; \sigma_n^2\;\text{Tr}\left((\mathbf{B}_\mathcal{S}^\top\mathbf{B}_\mathcal{S})^{-1}\right), \qquad \|\mathbf{x}\|_0 = K, \tag{5.7}$$

where 'Tr' is the trace-operator. When the columns of $\mathbf{B}_\mathcal{S}$ are orthogonal, the constrained CRB takes the smallest value.

The proof of this statement is taken from [30][5] and given below.

[5] C. Weiss and A. M. Zoubir, "A Compressed Sampling and Dictionary Learning Framework for Wavelength-Division-Multiplexing-Based Distributed Fiber Sensing," accepted for publication in *Journal of the Optical Society of America A*, 2017 (assigned issue: vol. 34, no. 5).

*Proof:*

The eigenvalues of the positive semi-definite matrix $\mathbf{H} = \mathbf{B}_{\mathcal{S}}^{\top}\mathbf{B}_{\mathcal{S}}$ are denoted by $\lambda_k \geq 0,\ k = 1, \ldots, K$. Then, using 'Hadamard's inequality' (c.f. [142]),

$$\det(\mathbf{H}) = \prod_{k=1}^{K} \lambda_k \ \leq \ \prod_{k=1}^{K} h_{kk} = \prod_{k=1}^{K} \widetilde{\lambda}_k = \det(\,\mathrm{diag}(\widetilde{\lambda}_1, \ldots, \widetilde{\lambda}_K)\,), \tag{5.8}$$

where $h_{kk}$ are the diagonal entries of $\mathbf{H}$ and $\widetilde{\lambda}_k$ are the eigenvalues of some diagonal matrix, $\mathrm{diag}(\widetilde{\lambda}_1, \ldots, \widetilde{\lambda}_K)$. The inverse of the matrix $\mathbf{H}$ fulfills the equation

$$\det(\mathbf{H}^{-1}) = \prod_{k=1}^{K} \frac{1}{\lambda_k} \ \geq \ \prod_{k=1}^{K} \frac{1}{\widetilde{\lambda}_k} = \det(\,[\mathrm{diag}(\widetilde{\lambda}_1, \ldots, \widetilde{\lambda}_K)]^{-1}\,)\,. \tag{5.9}$$

Therefore, a lower bound is obtained when $\mathbf{H}$ is a diagonal, such that equality holds in (5.8) and (5.9). This happens when the columns of $\mathbf{B}$ are orthogonal. ■

Usually, the columns of $\mathbf{B}$ are normalized, i.e. $\|\mathbf{b}_k\|_2^2 = \widetilde{\lambda}_k = 1$. Then, a lower bound for the MSE is found by

$$\mathrm{MSE}(\hat{\mathbf{x}}, \mathbf{x}) \ \geq \ \sigma_n^2 \sum_{k=1}^{K} \frac{1}{\widetilde{\lambda}_k} \ = \ K\,\sigma_n^2, \qquad \|\mathbf{x}\|_0 = K. \tag{5.10}$$

### 5.4.2 Estimating a Dictionary Parameter

In this subsection, the CRB for estimating $\theta$ is derived. The sparse coefficients, $\mathbf{x}$, are assumed to be known.

The derivation below is partly taken from [131][6].

At first, the Fisher information, $\mathcal{I}(\theta)$, is determined. To this end, the first and second moments of the score function, $\boldsymbol{\Delta}_{\mathrm{sc}}^{(\theta)} : \mathbb{R} \to \mathbb{R}$, are determined. It is given by the partial derivative of the log-likelihood function in (5.3) with respect to the parameter of interest, $\theta$. For i.i.d. Gaussian measurements, $y_m,\ m = 1, \ldots, M$, one obtains

$$\begin{aligned}
\boldsymbol{\Delta}_{\mathrm{sc}}^{(\theta)} &= -\frac{1}{2\sigma_n^2}\frac{\partial}{\partial\theta}\left(\mathbf{y}^{\top}\mathbf{y} - \mathbf{x}^{\top}\mathbf{D}(\theta)^{\top}\boldsymbol{\Phi}^{\top}\mathbf{y} - \mathbf{y}^{\top}\boldsymbol{\Phi}\mathbf{D}(\theta)\mathbf{x} + \mathbf{x}^{\top}\mathbf{D}(\theta)^{\top}\boldsymbol{\Phi}^{\top}\boldsymbol{\Phi}\mathbf{D}(\theta)\mathbf{x}\right) \\
&= -\frac{1}{2\sigma_n^2}\left(-2\mathbf{x}^{\top}\mathbf{D}'(\theta)^{\top}\boldsymbol{\Phi}^{\top}\mathbf{y} + \mathbf{x}^{\top}\mathbf{D}'(\theta)^{\top}\boldsymbol{\Phi}^{\top}\boldsymbol{\Phi}\mathbf{D}(\theta)\mathbf{x} + \mathbf{x}^{\top}\mathbf{D}(\theta)^{\top}\boldsymbol{\Phi}^{\top}\boldsymbol{\Phi}\mathbf{D}'(\theta)\mathbf{x}\right) \\
&= -\frac{1}{2\sigma_n^2}\left(-2\mathbf{x}^{\top}\mathbf{D}'(\theta)^{\top}\boldsymbol{\Phi}^{\top}[\,\mathbf{y} - \boldsymbol{\Phi}\mathbf{D}(\theta)\mathbf{x}\,]\right) \\
&= \frac{1}{\sigma_n^2}\left(\mathbf{x}^{\top}\mathbf{D}'(\theta)^{\top}\boldsymbol{\Phi}^{\top}[\,\mathbf{y} - \boldsymbol{\Phi}\mathbf{D}(\theta)\mathbf{x}\,]\right),
\end{aligned} \tag{5.11}$$

[6] C. Weiss and A. M. Zoubir, "Dictionary Learning Strategies for Compressed Fiber Sensing Using a Probabilistic Sparse Model," submitted to *IEEE Transactions on Signal Processing, 2016.*

where $\mathbf{D}'(\theta)$ is the element-wise derivative of $\mathbf{D}(\theta)$ with respect to $\theta$. Let $p(\mathbf{y}\,|\,\mathbf{x},\theta)$ denote the probability density of $\mathbf{y}$ given $\mathbf{x}$ and $\theta$, and let $\mathbb{E}_{\mathbf{y}|\mathbf{x},\theta}$ denote the corresponding conditional expectation operator. Then, $\mathbb{E}_{\mathbf{y}|\mathbf{x},\theta}\ \mathbf{y} = \mathbf{\Phi}\mathbf{D}(\theta)\mathbf{x}$ and the first moment of the score function vanishes, since

$$\begin{aligned}\mathbb{E}_{\mathbf{y}|\mathbf{x},\theta}\ \boldsymbol{\Delta}_{\text{sc}}^{(\theta)} &= \int \frac{1}{\sigma_n^2}\,(\,\mathbf{y}-\mathbf{\Phi}\mathbf{D}(\theta)\mathbf{x}\,)^\top\,\mathbf{\Phi}\,\mathbf{D}(\theta)\ p(\mathbf{y}\,|\,\mathbf{x},\theta)\,d\mathbf{y}\\ &= \frac{1}{\sigma_n^2}\mathbf{D}(\theta)^\top\mathbf{\Phi}^\top\left[\,\left(\mathbb{E}_{\mathbf{y}|\mathbf{x},\theta}\ \mathbf{y}\,\right)\ -\ \mathbf{\Phi}\mathbf{D}(\theta)\mathbf{x}\,\right]\ =\ \mathbf{0}\,. \qquad (5.12)\end{aligned}$$

The Fisher information is defined as the second moment of the score function, i.e. $\mathbb{E}_{\mathbf{y}|\mathbf{x},\theta}\ (\boldsymbol{\Delta}_{\text{sc}}^{(\theta)})^2$. Under certain regularity conditions (see Appendix A.2) and using (5.12), it can be calculated by

$$\begin{aligned}\mathcal{I}(\theta) &= \mathbb{E}_{\mathbf{y}|\mathbf{x},\theta}\ \frac{\partial^2}{\partial\theta^2}\log p(\mathbf{y}\,|\,\mathbf{x},\theta)\ =\ \mathbb{E}_{\mathbf{y}|\mathbf{x},\theta}\ \frac{\partial}{\partial\theta}\boldsymbol{\Delta}_{\text{sc}}^{(\theta)}\\ &= \mathbb{E}_{\mathbf{y}|\mathbf{x},\theta}\ \frac{1}{\sigma_n^2}\left(\mathbf{x}^\top\mathbf{D}''(\theta)^\top\mathbf{\Phi}^\top\mathbf{y}-\mathbf{x}^\top\mathbf{D}''(\theta)^\top\mathbf{\Phi}^\top\mathbf{\Phi}\mathbf{D}(\theta)\mathbf{x}+\mathbf{x}^\top\mathbf{D}'(\theta)^\top\mathbf{\Phi}^\top\mathbf{\Phi}\mathbf{D}'(\theta)\mathbf{x}\right)\\ &= \frac{1}{\sigma_n^2}\left(\mathbf{x}^\top\mathbf{D}''(\theta)^\top\mathbf{\Phi}^\top\mathbf{\Phi}\mathbf{D}(\theta)\mathbf{x}-\mathbf{x}^\top\mathbf{D}''(\theta)^\top\mathbf{\Phi}^\top\mathbf{\Phi}\mathbf{D}(\theta)\mathbf{x}+\mathbf{x}^\top\mathbf{D}'(\theta)^\top\mathbf{\Phi}^\top\mathbf{\Phi}\mathbf{D}'(\theta)\mathbf{x}\right)\\ &= \frac{1}{\sigma_n^2}\,\mathbf{x}^\top\mathbf{D}'(\theta)^\top\mathbf{\Phi}^\top\mathbf{\Phi}\mathbf{D}'(\theta)\mathbf{x}, \qquad (5.13)\end{aligned}$$

where $\mathbf{D}''(\theta)$ is the second (element-wise) derivative of $\mathbf{D}(\theta)$ with respect to $\theta$. Finally, the variance of any unbiased estimator, $\hat{\theta}$, for estimating some parameter $\theta\in\Theta$ given $\mathbf{x}$, is lower bounded by

$$\text{Var}(\hat{\theta})\ \leq\ \mathcal{I}^{-1}(\theta)\ =\ \frac{\sigma_n^2}{\mathbf{x}^\top\mathbf{D}'(\theta)^\top\mathbf{\Phi}^\top\mathbf{\Phi}\mathbf{D}'(\theta)\mathbf{x}}, \qquad \theta\in\Theta. \qquad (5.14)$$

Due to the unbiasedness, Equation (5.14) also represents a lower bound for the MSE of $\hat{\theta}$, i.e. $\text{MSE}(\hat{\theta},\theta)=\text{Var}(\hat{\theta})$.

Yet, the obtained bounds for $\text{MSE}(\hat{\mathbf{x}},\mathbf{x})$ and $\text{MSE}(\hat{\theta},\theta)$ do not consider the mutual information shared between $\mathbf{x}$ and $\theta$. Therefore, the CRB for jointly estimating $(\mathbf{x},\theta)$ is derived in the next subsection.

### 5.4.3 Estimating a Sparse Representation and a Dictionary Parameter

The previously obtained bounds are valid for estimating either $\mathbf{x}$ or $\theta$, given full knowledge of the other variable. They are based on the assumption that knowledge of $\mathbf{x}$ does not reveal information of $\theta$ or vice versa. However, $\mathbf{x}$ and $\theta$ share some mutual

information, such that the combined FIM is not diagonal.

The derivation below is partly taken from [131][7].

In order to incorporate an unknown parameter, $\theta$, into the framework of the constrained CRB for $\mathbf{x}$, the combined parameter tuple, $\boldsymbol{\gamma} = (\mathbf{x}^\top, \theta) \in \{\mathbb{R}_+^N \times \Theta\}$, is considered. The full FIM takes the form

$$\boldsymbol{\mathcal{I}}(\boldsymbol{\gamma}) = \begin{pmatrix} \boldsymbol{\mathcal{I}}(\mathbf{x}) & \breve{\mathbf{u}} \\ \breve{\mathbf{v}}^\top & \mathcal{I}(\theta) \end{pmatrix}, \tag{5.15}$$

where

$$\breve{\mathbf{u}} = -\mathbb{E}_{\mathbf{y}|\mathbf{x},\theta} \left[ \frac{\partial^2}{\partial x_1 \partial \theta} \log p(\mathbf{y}\,|\,\mathbf{x},\theta) \;,\; \ldots \;,\; \frac{\partial^2}{\partial x_N \partial \theta} \log p(\mathbf{y}\,|\,\mathbf{x},\theta) \right]^\top, \tag{5.16}$$

$$\breve{\mathbf{v}} = -\mathbb{E}_{\mathbf{y}|\mathbf{x},\theta} \left[ \frac{\partial^2}{\partial \theta \partial x_1} \log p(\mathbf{y}\,|\,\mathbf{x},\theta) \;,\; \ldots \;,\; \frac{\partial^2}{\partial \theta \partial x_N} \log p(\mathbf{y}\,|\,\mathbf{x},\theta) \right]^\top. \tag{5.17}$$

The $i$-th element of $\breve{\mathbf{u}}$, $i = 1, \ldots, N$, can be calculated by

$$\begin{aligned} & -\mathbb{E}_{\mathbf{y}|\mathbf{x},\theta} \, \frac{\partial^2}{\partial x_i \partial \theta} \log p(\mathbf{y}\,|\,\mathbf{x},\theta) \\ = \; & -\mathbb{E}_{\mathbf{y}|\mathbf{x},\theta} \, \frac{1}{\sigma_n^2} \frac{\partial}{\partial x_i} \left( \mathbf{x}^\top \mathbf{D}'(\theta)^\top \boldsymbol{\Phi}^\top [\, \mathbf{y} - \boldsymbol{\Phi}\mathbf{D}(\theta)\mathbf{x} \,] \right) \\ = \; & -\frac{1}{\sigma_n^2} \, \mathbb{E}_{\mathbf{y}|\mathbf{x},\theta} \left( \, [\mathbf{d}_i'(\theta)]^\top \boldsymbol{\Phi}^\top \mathbf{y} \, - \, [\mathbf{d}_i'(\theta)]^\top \boldsymbol{\Phi}^\top \boldsymbol{\Phi} \mathbf{D}(\theta)\mathbf{x} \, - \, \mathbf{x}^\top \mathbf{D}'(\theta)^\top \boldsymbol{\Phi}^\top \boldsymbol{\Phi} \, \mathbf{d}_i(\theta) \, \right) \\ = \; & \frac{1}{\sigma_n^2} \mathbf{x}^\top \mathbf{D}'(\theta)^\top \boldsymbol{\Phi}^\top \boldsymbol{\Phi} \, \mathbf{d}_i(\theta). \end{aligned} \tag{5.18}$$

[7] C. Weiss and A. M. Zoubir, "Dictionary Learning Strategies for Compressed Fiber Sensing Using a Probabilistic Sparse Model," submitted to *IEEE Transactions on Signal Processing, 2016.*

By symmetry, exchanging the order of the partial derivatives yields the same result for the $i$-th element of $\breve{\mathbf{v}}$, i.e

$$\begin{aligned}
& -\mathbb{E}_{\mathbf{y}|\mathbf{x},\theta}\, \frac{\partial^2}{\partial\theta\partial x_i} \log\, p(\mathbf{y}\,|\,\mathbf{x},\theta) \\
= \;& \mathbb{E}_{\mathbf{y}|\mathbf{x},\theta}\, \frac{\partial^2}{\partial\theta\partial x_i}\, \frac{1}{2\sigma_n^2}\, \big(\mathbf{y}^\top\mathbf{y} - 2\mathbf{y}^\top\boldsymbol{\Phi}\,\mathbf{D}(\theta)\mathbf{x} + \mathbf{x}^\top\mathbf{D}(\theta)^\top\boldsymbol{\Phi}^\top\boldsymbol{\Phi}\,\mathbf{D}(\theta)\mathbf{x}\,\big) \\
= \;& \mathbb{E}_{\mathbf{y}|\mathbf{x},\theta}\, \frac{\partial}{\partial\theta}\, \frac{1}{2\sigma_n^2}\, \big(-2\,\mathbf{y}^\top\boldsymbol{\Phi}\,\mathbf{d}_i(\theta) + [\mathbf{d}_i(\theta)]^\top\boldsymbol{\Phi}^\top\boldsymbol{\Phi}\,\mathbf{D}(\theta)\mathbf{x} + \mathbf{x}^\top\mathbf{D}(\theta)^\top\boldsymbol{\Phi}^\top\boldsymbol{\Phi}\,\mathbf{d}_i(\theta)\,\big) \\
= \;& \mathbb{E}_{\mathbf{y}|\mathbf{x},\theta}\, \frac{1}{2\sigma_n^2}\, \big(-2\,\mathbf{y}^\top\boldsymbol{\Phi}\,\mathbf{d}_i'(\theta) + \mathbf{d}_i'(\theta)^\top\boldsymbol{\Phi}^\top\boldsymbol{\Phi}\,\mathbf{D}(\theta)\mathbf{x} + [\mathbf{d}_i(\theta)]^\top\boldsymbol{\Phi}^\top\boldsymbol{\Phi}\,\mathbf{D}'(\theta)\mathbf{x} \\
& \qquad\qquad + \mathbf{x}^\top\mathbf{D}'(\theta)^\top\boldsymbol{\Phi}^\top\boldsymbol{\Phi}\,\mathbf{d}_i(\theta) + \mathbf{x}^\top\mathbf{D}(\theta)^\top\boldsymbol{\Phi}^\top\boldsymbol{\Phi}\,\mathbf{d}_i'(\theta)\,\big) \\
= \;& \frac{1}{2\sigma_n^2}\, \big(-2\,\mathbf{x}^\top\mathbf{D}(\theta)^\top\boldsymbol{\Phi}^\top\boldsymbol{\Phi}\,\mathbf{d}_i'(\theta) + 2\,\mathbf{x}^\top\mathbf{D}(\theta)^\top\boldsymbol{\Phi}^\top\boldsymbol{\Phi}\,\mathbf{d}_i'(\theta) \\
& \qquad\qquad + [\mathbf{d}_i(\theta)]^\top\boldsymbol{\Phi}^\top\boldsymbol{\Phi}\,\mathbf{D}'(\theta)\mathbf{x} + \mathbf{x}^\top\mathbf{D}'(\theta)^\top\boldsymbol{\Phi}^\top\boldsymbol{\Phi}\,\mathbf{d}_i(\theta)\,\big) \\
= \;& \frac{1}{\sigma_n^2}\mathbf{x}^\top\mathbf{D}'(\theta)^\top\boldsymbol{\Phi}^\top\boldsymbol{\Phi}\,\mathbf{d}_i(\theta)\,. \qquad (5.19)
\end{aligned}$$

Therefore, one obtains

$$\begin{aligned}
\mathbb{E}_{\mathbf{y}|\mathbf{x},\theta}\, \breve{\mathbf{u}} \;&=\; \mathbb{E}_{\mathbf{y}|\mathbf{x},\theta}\, \breve{\mathbf{v}} \;=\; -\frac{1}{\sigma_n^2}\, \big(\,\mathbf{x}^\top\mathbf{D}'(\theta)^\top\boldsymbol{\Phi}^\top\boldsymbol{\Phi}\,\mathbf{D}(\theta)\,\big)^\top \\
&=\; -\frac{1}{\sigma_n^2}\mathbf{D}(\theta)\,\boldsymbol{\Phi}^\top\boldsymbol{\Phi}\,\mathbf{D}'(\theta)\,\mathbf{x}\,. \qquad (5.20)
\end{aligned}$$

In order to extend the notation of the constrained CRB for the tuple $\boldsymbol{\gamma} = (\mathbf{x},\theta)$, the set of feasible directions can be extended by one additional direction at index $(N{+}1)$, corresponding to $\theta$. For any $\theta \in \Theta$, this direction is always feasible as it is not constrained by the sparsity assumption for $\mathbf{x}$. The extended projection matrix becomes $\widetilde{\mathbf{U}}_f = [\,\mathbf{e}_{i_1},\ldots,\mathbf{e}_{i_K},\mathbf{e}_{N+1}\,]$ and the reduced FIM is obtained by

$$\boldsymbol{\mathcal{I}}_{K+1}(\boldsymbol{\gamma}) \;=\; \widetilde{\mathbf{U}}_f^\top\, \boldsymbol{\mathcal{I}}(\mathbf{x},\theta)\, \widetilde{\mathbf{U}}_f \qquad (5.21)$$

In order to find the inverse of this matrix, the matrix inversion lemma in block form [143] can be applied, i.e.

$$\boldsymbol{\mathcal{I}}_{K+1}^{-1}(\boldsymbol{\gamma}) \;=\; \begin{pmatrix} \Big(\boldsymbol{\mathcal{I}}_K(\mathbf{x}) - \frac{1}{\mathcal{I}(\theta)}\breve{\mathbf{v}}\,\breve{\mathbf{v}}^\top\Big)^{-1} & -\frac{1}{\breve{b}}\,\boldsymbol{\mathcal{I}}_K^{-1}(\mathbf{x})\,\breve{\mathbf{v}} \\ -\frac{1}{\breve{b}}\,\breve{\mathbf{v}}^\top\,\boldsymbol{\mathcal{I}}_K^{-1}(\mathbf{x}) & \frac{1}{\breve{b}} \end{pmatrix}, \qquad (5.22)$$

with $\breve{b} \;=\; \mathcal{I}(\theta) \;-\; \breve{\mathbf{v}}^\top\,\boldsymbol{\mathcal{I}}_K^{-1}(\mathbf{x})\,\breve{\mathbf{v}}$, and

$$\Big(\boldsymbol{\mathcal{I}}_K(\mathbf{x}) - \frac{1}{\mathcal{I}(\theta)}\breve{\mathbf{v}}\,\breve{\mathbf{v}}^\top\Big)^{-1} \;=\; \boldsymbol{\mathcal{I}}_K^{-1}(\mathbf{x}) \;+\; \frac{1}{\breve{b}}\,\boldsymbol{\mathcal{I}}_K^{-1}(\mathbf{x})\,\breve{\mathbf{v}}\,\breve{\mathbf{v}}^\top\,\boldsymbol{\mathcal{I}}_K^{-1}(\mathbf{x})\,. \qquad (5.23)$$

Consequently, the constrained CRB for an unbiased estimator $\hat{\boldsymbol{\gamma}} = (\hat{\mathbf{x}}, \hat{\theta})$ becomes

$$\mathrm{Cov}(\hat{\boldsymbol{\gamma}}) \;\succeq\; \widetilde{\mathbf{U}}_f \, \boldsymbol{\mathcal{I}}_{K+1}^{-1}(\boldsymbol{\gamma}) \, \widetilde{\mathbf{U}}_f^{\top}, \qquad \|\mathbf{x}\|_0 = K, \quad \theta \in \Theta. \tag{5.24}$$

Finally, the lower bounds for the MSE corresponding to the individual parameters can be determined from the diagonal elements of $\boldsymbol{\mathcal{I}}_{K+1}^{-1}(\boldsymbol{\gamma})$ in (5.22).

Compared to the case where $\mathbf{x}$ or $\theta$ are individually estimated using $\boldsymbol{\mathcal{I}}_K(\mathbf{x})$ or $\mathcal{I}(\theta)$, the lower bound of the respective MSE can be obtained by adding a "correction term" that accounts for the mutual information shared between these two variables, i.e.

$$\mathrm{MSE}(\hat{\mathbf{x}}, \mathbf{x}) \;\geq\; \mathrm{Tr}\left( \boldsymbol{\mathcal{I}}_K^{-1}(\mathbf{x}) \;+\; \frac{1}{\breve{b}} \boldsymbol{\mathcal{I}}_K^{-1}(\mathbf{x}) \, \breve{\mathbf{v}} \, \breve{\mathbf{v}}^{\top} \boldsymbol{\mathcal{I}}_K^{-1}(\mathbf{x}) \right) \tag{5.25}$$

$$= \left( \mathrm{Tr} \;\boldsymbol{\mathcal{I}}_K^{-1}(\mathbf{x}) \right) \;+\; \underbrace{\frac{1}{\breve{b}} \, \breve{\mathbf{v}}^{\top} \boldsymbol{\mathcal{I}}_K^{-1}(\mathbf{x}) \, \boldsymbol{\mathcal{I}}_K^{-1}(\mathbf{x}) \, \breve{\mathbf{v}}}_{\text{“correction”}} \tag{5.26}$$

and

$$\mathrm{MSE}(\hat{\theta}, \theta) \;\geq\; \frac{1}{\breve{b}} \;=\; \mathcal{I}(\theta)^{-1} \;+\; \underbrace{\frac{\breve{\mathbf{v}}^{\top} \boldsymbol{\mathcal{I}}_K^{-1} \, \breve{\mathbf{v}}}{\mathcal{I}(\theta) \left[ \mathcal{I}(\theta) \;-\; \breve{\mathbf{v}}^{\top} \boldsymbol{\mathcal{I}}_K^{-1} \, \breve{\mathbf{v}} \right]}}_{\text{“correction”}} . \tag{5.27}$$

The existence of these bounds is limited to the constraint $\breve{b} = \mathcal{I}(\theta) - \breve{\mathbf{v}}^{\top} \boldsymbol{\mathcal{I}}_K^{-1}(\mathbf{x}) \, \breve{\mathbf{v}} \neq 0$.

## 5.5 A Unified Framework for Compressed Fiber Sensing With Uncertainty

In this section, a *compressed fiber sensing and dictionary learning* (CFS-DL) framework is presented. It provides a parametric model to describe the observed sensor signal, $\mathbf{r}$, in (5.2), for a basic system architecture. This model is used to create a parametric translation-invariant dictionary according to (5.1). The presented framework unifies CS-based acquisition, sparse estimation and DL for general parametric dictionaries with high coherence levels and uncertain parameters.

The material presented in this section is partly taken from [87][8], [88][9], [30][10].

### 5.5.1 System Architecture

The core architecture of the considered fiber sensing system is depicted in Figure 5.1. This system is setup is based on the fiber sensor presented in [70, 86] and extended by a CS acquisition module and a signal processing block for sparse estimation and DL. The same setup is considered subsequently in Section 5.7 for simulations and for experimental validation using real data from the fiber sensor in [70, 86].
It is assumed that the system architecture has a modular structure based on *linear time-invariant* (LTI) components, where additional LTI components may be added. According to Figure 5.1, the emitted laser signal passes a *single-mode fiber* (SMF) before entering the sensing fiber. When the signal wavelength falls inside the reflection spectrum of an FBG, it is reflected back and redirected to the receiver. The optical power is directly detected by a *photodetector* (PD). The electrical output signal is acquired using CS. Finally, the desired information is extracted in the signal processing block using the PDL-OIAI/OMP algorithms for sparse estimation and DL, introduced in Section 5.5.4.

[8] C. Weiss and A. M. Zoubir, "Fiber Sensing Using UFWT-Lasers and Sparse Acquisition," in *Proc. of the 21st European Signal Processing Conference (EUSIPCO)*, September, 2013.

[9] C. Weiss and A. M. Zoubir, "Fiber Sensing Using Wavelength-Swept Lasers: A Compressed Sampling Approach," in *Proc. of the 3rd International Workshop on Compressed Sensing Theory and its Applications to Radar, Sonar and Remote Sensing (CoSeRa)*, June, 2015.

[10] C. Weiss and A. M. Zoubir, "A Compressed Sampling and Dictionary Learning Framework for Wavelength-Division-Multiplexing-Based Distributed Fiber Sensing," accepted for publication in *Journal of the Optical Society of America A*, 2017 (assigned issue: vol. 34, no. 5).

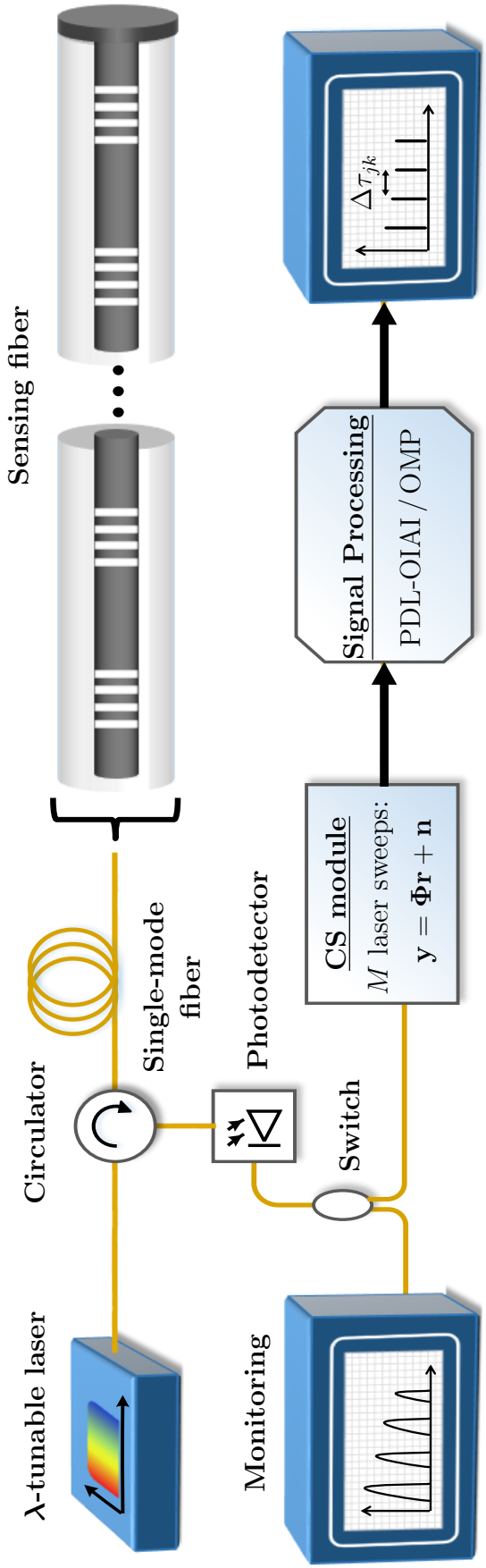



Figure 5.1. System layout for WDM-based quasi-distributed fiber-optic sensing with CS-based acquisition and post-processing.

### 5.5.2 A Parametric Signal Model for Fiber-optic Sensing

In order to obtain a parametric signal model for the observed sensor signal, $\mathbf{r}$, different existing physical models are compiled. Due to the modular architecture, each subsystem can be described separately. The subsystems between the laser and the photodetector are treated as LTI components with a well-defined input/output relation. Although the considered model is specified for the sensor architecture in Figure 5.1, it can be customized to describe various system configurations as long as the LTI assumption holds. The models for the individual components are detailed below.

#### 5.5.2.1 Laser Output:

Fiber interrogation is performed using a wavelength-tunable laser as in [70, 86]. It is an actively mode-locked fiber laser, which yields a pulsed output signal. A wideband *semiconductor optical amplifier* (SOA) is used as a gain medium and a *dispersion-compensating fiber* (DCF) forms the laser cavity. The mode-locking condition states, that the frequency of the SOA injection current, $f_m$, has to match the mode spacing of the laser cavity, which is given by the free spectral range. However, since the cavity is highly dispersive, the free spectral range depends on the lasing frequency and this condition is only fulfilled within a small spectral region. Thus, a change in $f_m$ also causes a shift in the frequency region where mode-locking is achieved. As a result the lasing frequency can be swept by changing $f_m$, which is referred to as 'dispersion tuning' [86].
When the wavelength is swept, each laser pulse is generated by the superimposed cavity modes around the instantaneous center wavelength, $\lambda_i$, where $i$ is the pulse index. The instantaneous pulse repetition rate can be described by the actual modulation frequency, $\tau_{\text{rep},i} = 1/f_{m,i}$. Thus, the output signal is a non-uniform pulse train with varying pulse repetition rate and pulse width.
This kind of lasers has several advantages for quasi-distributed fiber sensing [70, 86]. First, the absence of mechanical components, such as tunable filters in the laser cavity, reduces hardware costs and allows for high sweep rates. Thus, also time-varying perturbations can be monitored. Second, their wide tuning range can be used to interrogate a large number of FBGs, which enables sensing over long distances.
at high precision.
In the baseband frequency domain, the emitted laser pules of an actively mode-locked fiber laser with center wavelength $\lambda$ can be described by chirped Gaussian pulses, i.e. [70, 86, 144]

$$A^{(\lambda)}(\omega) \;\sim\; \exp\left(-\frac{\omega^2}{2\,(\delta\omega^{(\lambda)})^2}\right), \tag{5.28}$$

where the pulse bandwidth, $\delta\omega^{(\lambda)}$, is given by [70]

$$\delta\omega^{(\lambda)} \;=\; \sqrt{\pi\frac{f_m}{\lambda}}\left(\frac{8\pi\, c_0\, \Gamma_a}{|D^{(\lambda)}_{\mathrm{DCF}}|\, L_{\mathrm{DCF}}}\right)^{\frac{1}{4}}. \tag{5.29}$$

Herein, $c_0$ is the speed of light, $L_{\mathrm{DCF}}$ is the length of the DCF, $\Gamma_a$ denotes the amplitude modulation index of the light within the DCF (laser cavity), and $D^{(\lambda)}_{\mathrm{DCF}}$ is the first order dispersion parameter. The gain profile of the SOA at wavelength $\lambda$ can be described by [145, 146]

$$g^{(\lambda)} \;=\; \frac{a_1(N_c - N_0) - a_2(\lambda - \lambda_{N_c})^2 + a_3(\lambda - \lambda_{N_c})^3}{1 + \epsilon_c P_{\mathrm{av}}}\,. \tag{5.30}$$

Herein, $P_{\mathrm{av}}$ denotes the average output power over the length of the SOA, $\epsilon_c$ is a compression factor, $N_c$ is the actual carrier density and $N_0$ is the carrier density at the transparency point (no-gain wavelength). For $N_c$ and $N_0$, the maximum gain wavelengths are denoted by $\lambda_N$ and $\lambda_0$, respectively. Further, $a_1$ is the coefficient of the carrier difference, $a_2$ scales the spectral width of the gain profile, and $a_3$ accounts for any asymmetry of the gain profile. A third-order fit to the measured SOA gain can be used to find the coefficients $a_1, a_2, a_3$. Some typical values are listed in [145, 146]. A steady state numerical SOA model can be found in [147].

#### 5.5.2.2 Fiber Transmission:

The laser pulses travel through an SMF to the FBGs. The pulses are reflected if the wavelength matches the reflection spectrum of an FBG Therefore, the travel distance, $L^{(\lambda)}_{\mathrm{SMF}}$, changes for different wavelengths. It is assumed that non-linear effects can be neglected and that chromatic dispersion is the dominating effect.
The transfer function of the fiber can be derived from a Taylor expansion of the propagation constant, i.e. $\beta^{(\lambda)}(\omega) = \sum_{j=0}^{\infty} \beta_j^{(\lambda)} (\Delta\omega)^j$, where $\beta_2{}^{(\lambda)}$ and $\beta_3{}^{(\lambda)}$, describe the first and second order dispersion, respectively. The fiber damping is denoted by $\alpha_d^{(\lambda)}$. Then, the baseband transfer function is given by [148–150]

$$H^{(\lambda)}\left(\omega, L^{(\lambda)}_{\mathrm{SMF}}\right) \;=\; \exp\left(-\left(\alpha_d^{(\lambda)} + j\frac{\beta_2^{(\lambda)}}{2}\omega^2 + j\frac{\beta_3^{(\lambda)}}{6}\omega^3\right) L^{(\lambda)}_{\mathrm{SMF}}\right). \tag{5.31}$$

where $\beta_2{}^{(\lambda)}$ and $\beta_3{}^{(\lambda)}$ are related to the dispersion, $D^{(\lambda)}_{\mathrm{SMF}}$, and to the dispersion slope, $S^{(\lambda)}_{\mathrm{SMF}} = (d/d\lambda) D^{(\lambda)}_{\mathrm{SMF}}$, respectively. In particular, [149]

$$\beta_2^{(\lambda)} \;=\; -D^{(\lambda)}_{\mathrm{SMF}}\,\frac{\lambda^2}{2\pi c_0} \tag{5.32}$$

$$\beta_3^{(\lambda)} \;=\; \left(S^{(\lambda)}_{\mathrm{SMF}}\,\frac{\lambda^3}{4\pi c_0} - \beta_2{}^{(\lambda)}\right)\frac{\lambda}{\pi c_0}\,. \tag{5.33}$$

The damping can be modeled by [79]

$$\alpha_d^{(\lambda)} = A_R \frac{1}{\lambda^4} + B_\alpha + C_\alpha^{(\lambda)} \ . \tag{5.34}$$

Herein, $A_R$ denotes the Rayleigh scattering coefficient, $B_\alpha$ stands for wavelength-independent losses, e.g. microbending or waveguide imperfections, and $C_\alpha^{(\lambda)}$ describes other wavelength-dependent losses such as $OH^-$ absorption peaks. Again, a least-squared fitting to the measured fiber damping can be used to determine these coefficients. Moreover, a detailed model for $C_\alpha^{(\lambda)}$ can be found in [79].

#### 5.5.2.3 FBG Reflection:

In the considered FBG model, the fiber axis is assumed to be in parallel to the $z$-axis of a Cartesian coordinate system. A uniform (non-chirped) FBG can be described by a periodic variation of the refractive index in the fiber core [93, 151], i.e.

$$n_c(z) \ = \ n_0 \ + \ \Delta n_c \cos\left(\frac{2\pi(z - z_0)}{\Lambda_{\text{FBG}}}\right), \qquad z \in [z_0, z_0 + L_{\text{G}}] \ , \tag{5.35}$$

where $z_0$ is the location of the FBG, $L_{\text{G}}$ is the grating length, $n_0$ is the average refractive index, and $\Delta n_c$ is the amplitude of the index variation. The spectrum of an FBG can be derived using coupled-mode theory [93, 151–153]. Herein, one derives the field amplitudes of a mode in $+z$-direction and an identical counter-propagating mode in $-z$-direction, $S(z)$, $R(z)$, respectively. Their coupling is based on the dielectric perturbation. When the grating is uniform, $S(z)$ and $R(z)$ are determined by solving two coupled 1st-order ordinary differential equations with constant coefficients [89]:

$$\frac{dR}{dz} \ = \ j\,\sigma_c R(z) \ + \ j\,\kappa_c S(z) \qquad \text{and} \qquad \frac{dS}{dz} \ = \ -j\,\sigma_c S(z) \ + \ j\,\kappa_c^* R(z), \tag{5.36}$$

where $j$ is the imaginary unit and $(\cdot)^*$ denotes the complex conjugate. The parameters $\sigma_c$ and $\kappa_c$ are proportional to the refractive index. They are called the DC-/AC-coupling coefficients. Using suitable boundary conditions, a closed-form solution for $S(z)$ and $R(z)$ can be obtained. The ratio between the field amplitudes at $z = z_0$ is defined as the (wavelength-dependent) reflection coefficient of the field amplitudes [89]:

$$\rho^{(\lambda)} \ = \ \left.\frac{S(z = z_0)}{R(z = z_0)}\right|_\lambda \ = \ \frac{-\kappa_c \sinh(\gamma_c L_{\text{G}})}{\sigma_c \sinh(\gamma_c L_{\text{G}}) + j\,\gamma_c \cosh(\gamma_c L_{\text{G}})}, \tag{5.37}$$

where $\gamma_c = (\kappa_c^2 - \sigma_c^2)^{1/2}$. Non-uniform gratings with apodization or chirp can be modeled by assuming that the grating is piece-wise uniform, Hence, the FBG is subdivided into $M_G$ uniform segments in $z$-direction. For each segment, the corresponding field amplitudes, $R_q, S_q,\ q = 0, \ldots, M_G$, have to be calculated. The amplitudes

at the beginning and at the end of the grating are related by a transition matrix, $\mathbf{F} = \mathbf{F}_{M_G} \cdot \mathbf{F}_{M_G-1} \cdots \mathbf{F}_1$, where $\mathbf{F}_q\,,\ q = 1, \ldots, M_G$, are the transition matrices of the individual segments. Hence,

$$\begin{pmatrix} R_{M_G} \\ S_{M_G} \end{pmatrix} \;=\; \mathbf{F} \begin{pmatrix} R_0 \\ S_0 \end{pmatrix}. \tag{5.38}$$

#### 5.5.2.4 Received Sensor Signal:

At the last stage, the reflected signal travels back to the receiver, where it is converted to the electrical domain by the photodetector. Since the signal passes the fiber twice, the fiber transfer function, $H^{(\lambda)}(\omega, L^{(\lambda)}_{\mathrm{SMF}})$, has to be applied twice as well. Let $\Omega_0 = c_0/\lambda_0$ be the optical frequency at the center of the sweep range. Then, for the $i$-th pulse, the instantaneous center frequency is given by $\Omega_i = \Omega_0 - \Delta\Omega_i$, where $\Delta\Omega_i = c_0(\lambda_i - \lambda_0)/(\lambda_i\lambda_0)$.

Combining the models in (5.28, 5.30, 5.31, 5.37), an expression for the signal reflected from an FBG at distance $z = L_z$ can be obtained in the baseband frequency domain:

$$E_r(\omega) \;=\; \sum_i g^{(\lambda_i)}\, \rho^{(\lambda_i)}\, A^{(\lambda_i)}(\omega - \Delta\Omega_i)\, H^{(\lambda_i)}(\omega - \Delta\Omega_i, L_z)^2\, e^{-j\omega\tau_{\mathrm{rep},i}}\,. \tag{5.39}$$

Taking the inverse Fourier transform, the time-domain signal, $E_r(t, \lambda(t))$, can be determined. The dependency on $\lambda(t)$ emphasizes the change of the wavelength with time. Next, the received signal is detected by the photodetector. It is assumed that the light intensity is uniformly distributed over the sensitive area of the PD, $A_{\mathrm{PD}}$. Based on the optical intensity, $I(t)$, the detected optical power at time instant $t$ can be calculated by [154]

$$\bar{P}_r(t) \;=\; \int_{A_{\mathrm{PD}}} I(t)\, dA_{\mathrm{PD}} \;=\; \left| \frac{E_r(t, \lambda(t))\, E_r^*(t, \lambda(t))}{Z_W} \right|, \tag{5.40}$$

where $Z_W$ is the wave impedance of the medium. Then, the output current of the photodetector is given by [154]

$$i_{\mathrm{PD}}(t) \;=\; R^{(\lambda)}_{\mathrm{PD}}\, \bar{P}_r(t) \;=\; \frac{q_e\, \lambda(t)\, \eta^{(\lambda(t))}_{\mathrm{PD}}}{h\, c_0} \left| \frac{E_r(t, \lambda(t))\, E_r^*(t, \lambda(t))}{Z_W} \right|\,, \tag{5.41}$$

where $h$ is Planck's constant, $q_e$ is the elementary charge, and $\eta^{(\lambda)}_{\mathrm{PD}}$ is the quantum efficiency of the active material.
When the photodetector has a slow response time, the envelope signal that modulates the pulse train can be extracted. The overall bandwidth of the photodetector and the

receiver circuitry is modeled by a lowpass filter with transfer function $H_{\mathrm{LP}}^{(\Delta f)}(\omega)$ and an effective receiver bandwidth, $\Delta f$, i.e.

$$r(t, \Delta f) \;=\; \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{j\omega t}\, H_{\mathrm{LP}}^{(\Delta f)}(\omega)\, i_{\mathrm{PD}}(\omega)\, d\omega \;, \tag{5.42}$$

where $i_{\mathrm{PD}}(\omega)$ is the Fourier transform of $i_{\mathrm{PD}}(t)$.

Subsequently imperfect knowledge of $\Delta f$ is assumed. Since $r(t, \Delta f)$ is the generating function of the translation-invariant dictionary in (5.1), this assumption translates to uncertainty in the dictionary in terms of a dictionary parameter, $\theta$, which depends on $\Delta f$. It is a global parameter, because it jointly affects the temporal width of all FBG reflections.

### 5.5.3 Alternating Sparse Estimation and Dictionary Learning With Highly Coherent Dictionaries

In order to estimate a set of dictionary parameters, $\boldsymbol{\theta}$, and a sparse representation, $\mathbf{x}$, AE algorithms can be used The basic concept of AE is described below:
First, an initial estimate of $\boldsymbol{\theta}$ is used to obtain an estimate of the desired sparse representation, $\mathbf{x}$. In subsequent iterations, the current estimate of $\boldsymbol{\theta}$ is obtained based on the previous estimate of $\mathbf{x}$. The current estimate of $\boldsymbol{\theta}$ is then used to yield an improved estimate of $\mathbf{x}$, etc. The flow diagram in Figure 5.2 visualizes this process. It is detailed in Section 5.5.4.
In this chapter, sparse estimation is performed using a greedy OMP algorithm [7], although other algorithms can be similarly employed. The reasons for this choice are threefold:

(i) OMP offers a simple and fast implementation. Usually, it has a lower computational complexity than optimization-based methods [25].

(ii) OMP requires an estimate of the sparsity level to determine a stopping rule. In FBG-based fiber-optic sensing, the sparsity level is equivalent to the number of FBGs, which is exactly known.

(iii) Greedy methods are more robust to off-grid problems than optimization-based methods: OMP iteratively estimates one atom by considering the maximum correlation between all atoms and the residual signal. Therefore, it is guaranteed to yield a $K$-sparse representation after $K$ iterations. Optimization-based methods, in contrast, are likely to simultaneously select two adjacent atoms with reduced amplitudes when the true value falls in between the two grid points.

The translation-invariant dictionary used in CFS-DL can be highly coherent, such that the RIP conditions for the considered sparse estimation algorithm do not hold. Moreover, sparse estimation is performed based on the compressed measurements, $\mathbf{y}$, using the combined sensing matrix, $\mathbf{B}(\boldsymbol{\theta}) = \boldsymbol{\Phi}\mathbf{D}(\boldsymbol{\theta})$, where $\boldsymbol{\Phi}$ is a sub-Gaussian sampling matrix. Since $\mathbf{B}(\boldsymbol{\theta})$ represents the projection of $\mathbf{D}(\boldsymbol{\theta})$ onto a lower-dimensional subspace, the coherence level of $\mathbf{B}(\boldsymbol{\theta})$ is even higher. Besides, the standard RIP results for CS matrices are no longer applicable when a redundant dictionary is used (see Appendix A.1). Hence, standard OMP fails if the dictionary coherence is too strong, since the required RIP conditions for unique sparse reconstruction are violated [29,32]. In order to alleviate this problem, the coherence level is reduced by applying a dictionary pre-processing routine for inter-atom interference mitigation, as proposed in [32]. When IAI mitigation is applied, the desired sparse representations can be iteratively estimated along with the unknown dictionary parameters using AE. Although the RIP conditions for OMP are more stringent than for $\ell_1$-minimization [25], dictionary pre-processing is necessary in both cases if the coherence level is too high. The coherence distance, introduced in Chapter 2, Equation (2.11), is subsequently used to assess the difficulty in estimating the desired sparse representation and to distinguish the coherence level of dictionaries with different parametrizations.

### 5.5.4 The PDL-OIAI/OMP algorithms

The PDL-OIAI algorithm performs iterative OMP-based sparse estimation and DL using highly coherent parametric dictionaries. The inter-atom-interference (IAI) mitigation method in [32] is incorporated as a sub-routine to handle strong coherence distance. The PDL-OMP algorithm represents a reference method without IAI mitigation. It is introduced to emphasize the importance of IAI mitigation for alternating sparse estimation and DL when the dictionary is highly coherent. The flow diagram in Figure 5.2 depicts the full processing chain of both algorithms, including CS-based signal acquisition. Input for both algorithms are the CS samples, $\mathbf{y}$, a maximum number of AE-iterations, $D$, an initial guess for the dictionary parameters, $\hat{\boldsymbol{\theta}}^{(0)}$, and the corresponding combined sensing matrix, $\mathbf{B}(\hat{\boldsymbol{\theta}}^{(0)}) = \boldsymbol{\Phi}\mathbf{D}(\hat{\boldsymbol{\theta}}^{(0)})$.
In each AE iteration, the IAI mitigation sub-routine yields a data-dependent modified sensing dictionary, $\mathbf{W}$. Since the dictionary parameters are fixed during the IAI mitigation process, the dependency of $\mathbf{D}, \mathbf{B}$ and $\mathbf{W}$ on $\boldsymbol{\theta}$ can be omitted. Figure 5.3 depicts the reduced dictionary coherence of $\mathbf{W}$ for different initial coherence levels of $\mathbf{D}$. The number of measurements, $M$, (i.e. the number of rows in $\mathbf{B}$) is chosen to be large, such that $\mathbf{B}$ exhibits a similar coherence distance as $\mathbf{D}$. It can be seen that the coherence distance, $d_c(\mathbf{W})$ in Figure 5.3, is much smaller than $d_c(\mathbf{B})$ and $d_c(\mathbf{D})$, which

significantly alleviates the problem of sparse estimation. One major advantage is that the physical interpretation of the sparse support is not altered by the IAI mitigation procedure. In particular, the sparse representation, $\hat{\mathbf{x}}$, obtained with respect to $\mathbf{W}$, can be used to reconstruct the original sensor signal by $\hat{\mathbf{r}} = \mathbf{D}\hat{\mathbf{x}}$. Hence, the significant elements in $\hat{\mathbf{x}}$, indicated by the indices in $\hat{\mathcal{S}} = \text{supp}\{\hat{\mathbf{x}}\}$, can be used to estimate the sparse representation with respect to $\mathbf{D}$.

Following the AE paradigm in Figure 5.2, an initial sparse representation, $\hat{\mathbf{x}}^{(0)}$, is estimated based on $\mathbf{W}$ and based on the initial dictionary parameter, $\hat{\boldsymbol{\theta}}^{(0)}$. Then, a locally optimal estimate of $\boldsymbol{\theta}$ is obtained by minimizing a residual, $\text{res}(\boldsymbol{\theta})$. In the $d$-th iteration, an improved sparse representation, $\hat{\mathbf{x}}^{(d+1)}$, is estimated based on the current parameter estimates, $\hat{\boldsymbol{\theta}}^{(d)}$. Then, $\hat{\mathbf{x}}^{(d+1)}$ is used to improve the estimates of the parameters, yielding $\hat{\boldsymbol{\theta}}^{(d+1)}$, and so on. A suitable stopping rule can be defined by a lower threshold for $\text{res}^{(d)}(\boldsymbol{\theta})$, or by a maximum number of AE-iterations, $D$. The PDL-OIAI and PDL-OMP algorithms stop after $d = D$ iterations or if $\text{res}^{(d)}(\hat{\boldsymbol{\theta}}^{(d)}) \leq \beta_{\text{res}} = (1 + \epsilon_r)P_n$, for $\epsilon_r > 0$. Herein, $P_n$ is an estimate of the total noise power. The algorithms return the AE-iteration index, $d = d^*$, for which the smallest residual is obtained. It is used to determine the corresponding outputs, $\hat{\mathbf{x}}^{(d^*)}, \hat{\boldsymbol{\theta}}^{(d^*)}$ and $\hat{\mathcal{S}}^{(d^*)} = \text{supp}\{\hat{\mathbf{x}}^{(d^*)}\}$.

In Figure 5.2 the residual is calculated based on the $\ell_2$-norm of the difference between the measurements, $\mathbf{y}$, and the model, using the current estimates of $\mathbf{x}$ and $\boldsymbol{\theta}$. However, other cost functions can be used, dependent on the application. Generally, AE-based estimation is only locally optimal. Nevertheless, when the parameter space forms a convex set and also the objective function of the residual is jointly convex in all parameters, convex optimization methods can be readily applied to yield a globally optimal solution. When the dictionary is not a simple function of the parameters, it can be derived for a set of discrete values, such that the residual is minimized by comparing its value for all considered parameter values. The range of parameters can be limited by physical restrictions or based on *a priori* available information of the system. For example, a Fisher-information-based criterion can be used to determine a grid of parameter points with constant inter-atom coherence [155].

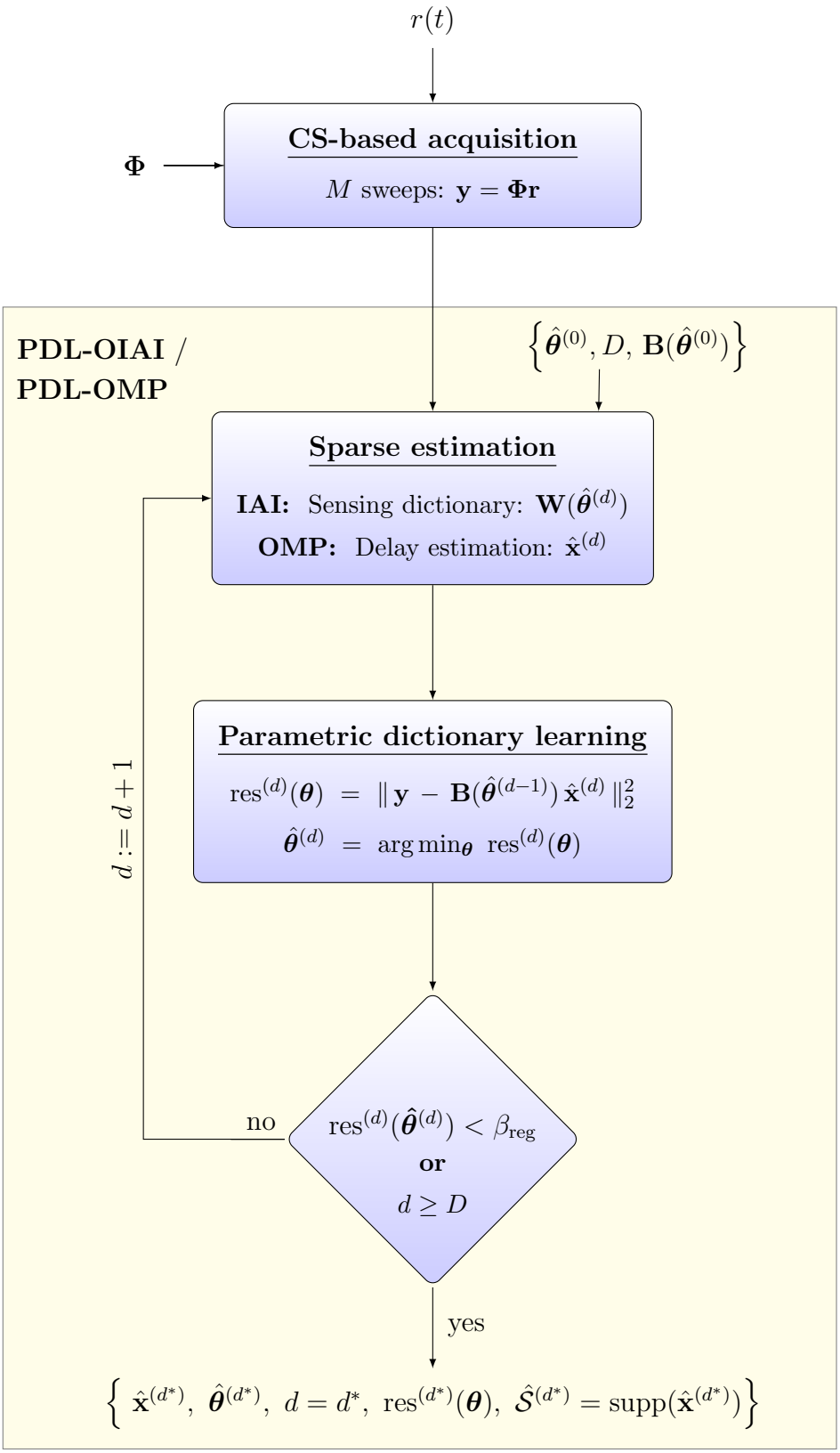



Figure 5.2. Structure diagram of the CFS-DL framework: CS-based acquisition is followed by PDL-OIAI/OMP for AM-based sparse estimation and parametric DL. The reference method, PDL-OMP, does not perform IAI mitigation.

#### 5.5.4.1 IAI mitigation

This subsection introduces the concept of IAI mitigation. It represents a short review of the method proposed in [32].

IAI mitigation alleviates the problem of high dictionary coherence, especially when greedy methods such as OMP are used. In contrast to related static preprocessing methods, e.g. in [29, 105], the data itself is taken into account to reduce the coherence level of the dictionary. Since the joint problem of CS with redundant dictionaries is considered, IAI mitigation is applied to the combined matrix, $\mathbf{B} = \mathbf{\Phi D}$.
In order to analyze the problem of IAI, it is advantageous to understand the underlying concept of the OMP algorithm. OMP pursues an iterative search for the dictionary atoms in $\mathbf{B} = [\mathbf{b}_1, \ldots, \mathbf{b}_N]$, based on their correlation with the measured signal, $\mathbf{y}$. It stops after a maximum number of iterations has been carried out. Let $\Omega = \{1, \ldots, N\}$ denote the set of dictionary indices, such that the index of the atom selected in the first ($k = 1$) OMP iteration is found by [32]

$$\arg\max_{i\in\Omega} |\mathbf{b}_i^\top \mathbf{y}| \;=\; \arg\max_{i\in\Omega} \left| \sum_{q\in\mathcal{S}} \mathbf{b}_i^\top \mathbf{b}_q x_q + \mathbf{b}_i^\top \mathbf{n} \right| . \tag{5.43}$$

The sum-term in (5.43) highlights the interference due to other atoms with indices in $\mathcal{S}$. Interference occurs when the inner product between different atoms is large, i.e. when $\mathbf{b}_i \mathbf{b}_q$ is large for $i \neq q$. This can lead to an incorrect selection of atoms and, hence, yield incorrect sparse representations. In order to reduce the coherence level, the problem (5.43) can be solved using a modified sensing dictionary, $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_N]$ [32], i.e.

$$\arg\max_{i\in\Omega} \left| \sum_{j\in\mathcal{S}} \mathbf{w}_i^\top \mathbf{b}_j x_j + \mathbf{w}_i^\top \mathbf{n} \right| . \tag{5.44}$$

The atoms of $\mathbf{W}$ have to be designed to minimize the IAI terms with other correct atoms, i.e. $\mathbf{w}_j^\top \mathbf{b}_i$, $j, i \in \mathcal{S}$, $i \neq j$. Nevertheless, in order to maintain the physical interpretation of the atom indices with respect to the original dictionary, $\mathbf{B}$ (or $\mathbf{D}$), strong correlation between $\mathbf{w}_i$ and the corresponding original atoms, $\mathbf{b}_i$, $i \in \mathcal{S}$, has to be enforced. Hence, when $\hat{\mathbf{x}}$ is a sparse representation of $\mathbf{y} = \mathbf{W}\mathbf{x}$, then $\hat{\mathbf{r}} = \mathbf{D}\hat{\mathbf{x}}$ is an approximation of the original signal, $\mathbf{r}$. The modified dictionary, $\mathbf{W}$, is found by calculating the Minimum Interference Distortionless Response [32]:

$$\mathbf{w}_i \;=\; \arg\min_{\tilde{\mathbf{w}}_i} \left| \, \tilde{\mathbf{w}}_i^\top \mathbf{B}_\mathcal{S} \, \mathbf{B}_\mathcal{S}^\top \, \tilde{\mathbf{w}}_i \right| \qquad \text{s.t.} \qquad \tilde{\mathbf{w}}_i^\top \mathbf{b}_i \;=\; 1 \,, \qquad i = 1, \ldots, N \,, \tag{5.45}$$

where $\mathbf{B}_\mathcal{S}$ is a sub-matrix of $\mathbf{B}$ that consists of the columns with indices in $\mathbf{S} = \text{supp}\{\mathbf{x}\}$. Since $\mathbf{S}$ is unknown, an iterative algorithm is used to approximate $\mathbf{B}_\mathcal{S}\mathbf{B}_\mathcal{S}^\top$ by

$\mathbf{B}\mathbf{U}^{(j)}\mathbf{B}^{\top}, j = 1, \dots J_U$, where $\mathbf{U}^{(j)} = \mathrm{diag}(|(\mathbf{W}^{(j-1)})^{\top}\mathbf{y}|^{\rho_u})$. Herein, $J_U$ denotes the number of iterations and $\rho_u > 0$ is a regularization parameter. Subsequent iterations with index $j$, $j = 1, \dots, J_U$, yield an improved solution for the modified dictionary atoms, $\mathbf{w}_i^{(j)}$, $i = 1, \dots N$, which can be calculated in closed-form according to [32].

Figure 5.3 shows the difference in the coherence levels of $\mathbf{W}$ as compared to the coherence of $\mathbf{D}$ and $\mathbf{B}$. The coherence is measured in terms of the coherence distance, $d_c$, introduced in Chapter 2, Equation 2.11. It can be seen that $d_c(\mathbf{W}) \ll d_c(\mathbf{B})$ and $d_c(\mathbf{W}) \ll d_c(\mathbf{D})$.

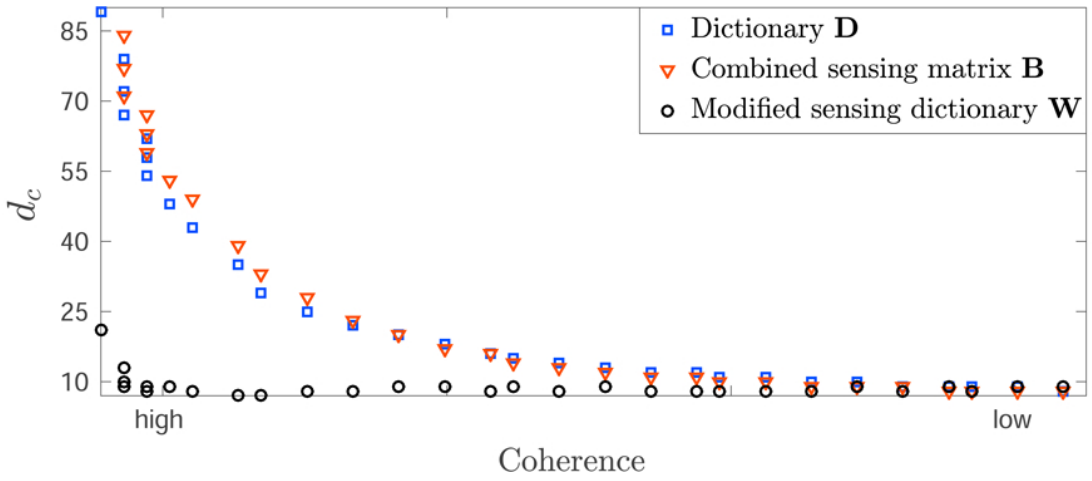



Figure 5.3. The coherence distance, $d_c$, as a measure of the coherence level for different matrices: The original coherent dictionary, $\mathbf{D}$, the combined sensing dictionary, $\mathbf{B}$, and the modified sensing dictionary, $\mathbf{W}$. The latter achieves the lowest values of the coherence distance due to IAI mitigation.

## 5.6 Computational Complexity

The material presented in this section is partly taken from [30][11].

The computational complexity is determined for both PDL-OIAI and PDL-OMP, assuming that $D$ AE iterations are performed. An AE iterations is subdivided into two steps, one for sparse estimation and the other for DL. In each AE iteration, both algorithms run $K$ iterations of the OMP algorithm to estimate the $K$ significant components in $\mathcal{S}$.
One OMP iteration has complexity $\mathcal{O}(NM)$, where $M$ is the number of CS measurements and $N$ is the dimensionality of the sparse coefficients. Hence, regarding the task of sparse estimation, PDL-OMP performs $\mathcal{O}(KMN)$ operations in the $d$-th AE iteration. For PDL-OIAI the task of sparse estimation is governed by IAI mitigation, i.e. by calculating the modified dictionary atoms $\mathbf{w}_i$ in (5.45) $\forall\ i = 1, \ldots, N$. The overall IAI mitigation procedure requires $J_U$ internal iterations to yield a modified sensing dictionary, $\mathbf{W}$, with sufficiently low coherence. Each internal iteration of the IAI mitigation sub-routine has complexity $\mathcal{O}(N^2M + NM^2 + M^3)$. Therefore, PDL-OIAI requires a total of $\mathcal{O}(KJ_U(N^2M + NM^2 + M^3))$ operations in the $d$-th AM iteration. Trading-off performance and computational complexity, $\mathbf{W}$ can be fixed after the first OMP iteration and calculated only for $k = 1$. In addition, the atoms, $\mathbf{w}_i,\ i = 1, \ldots, N$, can be independently calculated in parallel to speed up computations.
It is assumed that the dictionary is not a simple function of $\boldsymbol{\theta}$. Hence, it can only be calculated at discrete parameter points. Accordingly, the residual is calculated for a number of $\mathcal{R}_\theta$ values of the parameter $\boldsymbol{\theta}$. The complexity of computing the residual for one parameter point is $\mathcal{O}(NM^2)$. Finally, in the $d$-th AE iteration, the DL step requires $\mathcal{O}((R_\theta NM^2))$ operations for both PDL-OIAI and PDL-OMP.
The total computational complexity of PDL-OIAI and PDL-OMP is summarized in Table 5.1. For PDL-OIAI, the cases where $\mathbf{W}$ is fixed or re-calculated in each OMP iteration are explicitly distinguished.

Table 5.1. Computational complexity of PDL-OMP and of PDL-OIAI when $\mathbf{W}$ is determined $\forall\ k = 1, \ldots, K$, or only for $k = 1$, keeping it fixed for $k > 1$.

| | |
|---|---|
| **PDL-OMP**: | $\mathcal{O}(D[\,KNM + (R_\theta NM^2)\,])$ |
| **PDL-OIAI**: | $\mathcal{O}(D[\,KJ_U(N^3M + N^2M^2 + NM^3) + (R_\theta NM^2)\,])$ |
| **$\rightarrow$W fixed:** | $\mathcal{O}(D[J_U(N^3M+N^2M^2+NM^3)+KNM+(R_\theta NM^2)])$ |

[11] C. Weiss and A. M. Zoubir, "A Compressed Sampling and Dictionary Learning Framework for Wavelength-Division-Multiplexing-Based Distributed Fiber Sensing," accepted for publication in *Journal of the Optical Society of America A*, 2017 (assigned issue: vol. 34, no. 5).

## 5.7 Simulations and Experimental Validation

In this section, the applicability of the presented CFS-DL framework is evaluated based on simulations and real data. First, the parametric signal model in Section 5.5.2 is adopted. The parameters are adjusted to simulate the sensor system in [70, 86]. The parametrized model is used to generate the parametric dictionary in (5.1). Subsequently, this dictionary is used to evaluate the performance of PDL-OIAI/OMP for simulated data in various scenarios of different CS matrices, dictionary parameter values and SNRs. Finally, real data taken from the system in [70, 86] is used for experimental validation.

The experimental data was acquired at the Yamashita laboratory of photonic communication devices, The University of Tokyo, Japan. It was kindly provided by the authors Yamashita *et al.* in [86].
The material presented in this section is partly taken from [87][12], [88][13], [30][14].

### 5.7.1 Fiber-Optic Sensor Simulation

In order to create a suitable dictionary for sparse estimation, the sensor system in Figure 5.1 is simulated to obtain the received filtered sensor signal, $r(t, \Delta f)$, where $\Delta f$ is the filter bandwidth. This signal is used to generate the dictionary. Later on, the same dictionary is employed in simulations and for experimental validation using real data from the fiber sensor in [70, 86]. Therefore, the model parameters are adapted to match the system setup in [70, 86]. In this setup, the sensing fiber contains $K = 4$ FBGs, such that $K = 4$ reflections are observed at the receiver. Figure 5.4 (a) shows the total laser output (pulse train) of one sweep. The single pulses are closely spaced and not resolved in this plot. The shape of the envelope is mainly determined by the SOA gain profile. It is in good agreement with the experimental measurements in [70,86]. In Figure 5.4 (b), the FBG reflection spectrum is depicted. When the model parameters are correctly adjusted, the simulated reflection spectra are close to the reflection spectra of the FBGs in the sensing fiber in [70,86]. Figure 5.4 (c) depicts the

[12]C. Weiss and A. M. Zoubir, "Fiber Sensing Using UFWT-Lasers and Sparse Acquisition," in *Proc. of the 21st European Signal Processing Conference (EUSIPCO)*, September, 2013.

[13]C. Weiss and A. M. Zoubir, "Fiber Sensing Using Wavelength-Swept Lasers: A Compressed Sampling Approach," in *Proc. of the 3rd International Workshop on Compressed Sensing Theory and its Applications to Radar, Sonar and Remote Sensing (CoSeRa)*, June, 2015.

[14]C. Weiss and A. M. Zoubir, "A Compressed Sampling and Dictionary Learning Framework for Wavelength-Division-Multiplexing-Based Distributed Fiber Sensing," accepted for publication in *Journal of the Optical Society of America A*, 2017 (assigned issue: vol. 34, no. 5).

reflections, $r(t, \Delta f)$, from $K = 4$ FBGs. Herein, $\Delta f$ is chosen to extract the envelope signals without distorting their shape. The temporal width of the reflection envelopes is much smaller than the ones measured in [70, 86]. According to the authors in [70], the strong broadening of the experimentally measured reflections is caused by a limited bandwidth of the analog-to-digital converter and by an extended laser linewidth at high scan rates. Based on this observation, $\Delta f$ can be seen as an auxiliary parameter that accounts for all indistinguishable parameters that jointly contribute to the temporal width of the reflections. In order to estimate the underlying sparse representation, $\Delta f$ is estimated in terms of the global dictionary parameter, $\theta$.

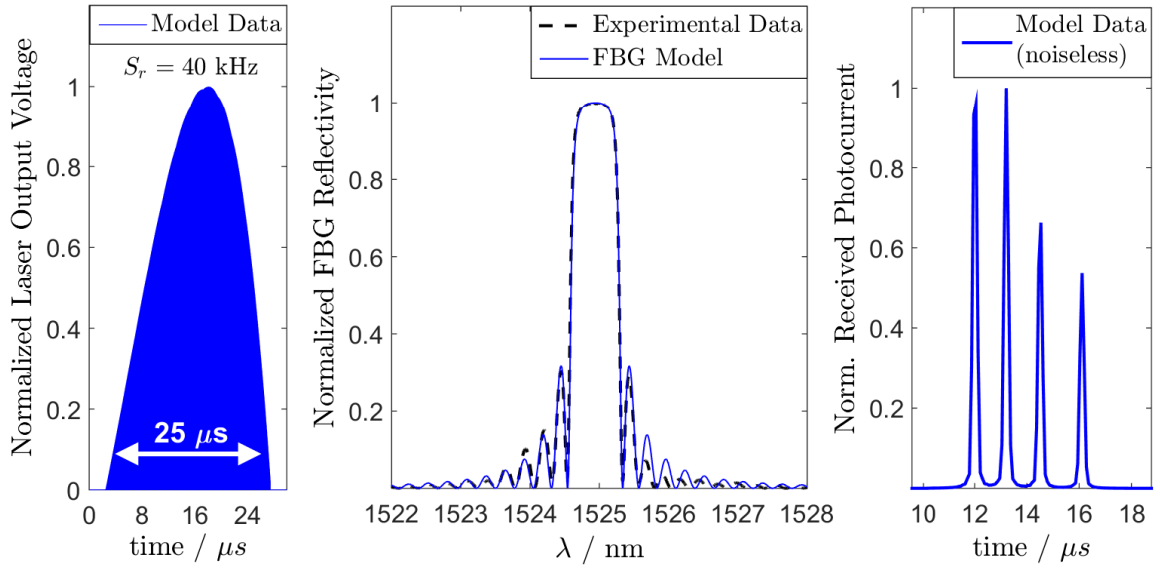



Figure 5.4. Fiber-optic sensor simulation. Left: laser output at sweep rate $S_r = 40$ kHz. Center: simulated and measured FBG reflectivity. Right: FBG reflection envelopes.

### 5.7.2 Setup and Basic Settings

In the subsequent simulations, $K = 3$ uniform FBGs are considered. Two reflections are closely spaced and all reflections have the same amplitude, $A_x$. The dictionary is based on the model in Section 5.5.2, where the $N = 2L$ dictionary atoms are created from $L = 134$ samples of $r(t, \Delta f)$. According to this model, the effective receiver bandwidth, $\Delta f$, determines the temporal width of the reflections. Since $K$ is small, the conventional coherence measures, $\mu_B(\mathbf{D}(\theta), K)$ and $\mu_{\text{MC}}(\mathbf{D}(\theta))$, are very similar and vary only slightly for small changes in the parameters $\delta t$, $\theta$. Therefore, the coherence distance, $d_c(\mathbf{D})$, introduced in Chapter 2, is used to distinguish between different coherence levels. The granularity of the delay grid is fixed to $\delta t = 50$ ns. Therefore,

a change in the coherence level corresponds to a change in $\Delta f$. It can be estimated in terms of a global dictionary parameter relative to the true value, yielding estimates $\hat{\theta} = \widehat{\Delta f}/\Delta f$. An initial estimate of $\theta$ is chosen at random, such that $\hat{\theta}^{(0)} \in [1.2, 5]$. The total parameter range is subdivided in $R_\theta \approx 100$ discrete parameter values.
CS is performed based on three types of sub-Gaussian sampling matrices, $\mathbf{\Phi}$. The entries of the matrices are i.i.d. and drawn from the distributions below:

$$\begin{aligned}
&\text{(I) Gauss :} && \mathcal{N}(0,1) \\
&\text{(II) Rademacher :} && \{\pm 1\} \quad \text{with equal probability} \\
&\text{(III) DF [106]:} && \{-1,0,1\} \text{ with probabilities } \{\tfrac{1}{6}, \tfrac{2}{3}, \tfrac{1}{6}\} \,.
\end{aligned}$$

The maximum number of AE iterations is set to $D = 8$. Based on the number of FBGs (i.e. the number of reflections), the PDL-OIAI/OMP algorithms perform $K = 3$ OMP iterations. For PDL-OIAI, the modified sensing dictionary, $\mathbf{W}$, is estimated using $J_U = 10$ iterations for IAI mitigation, where $\mathbf{W}$ is fixed after the first OMP iteration.
The *root mean-squared error* (RMSE) is considered to evaluate the estimation performance. Given a vector $\mathbf{v}$ and an estimator $\hat{\mathbf{v}}$, the RMSE is defined by

$$\mathrm{RMSE}(\mathbf{v}, \hat{\mathbf{v}}) = \sqrt{\mathbb{E}_{\hat{\mathbf{v}}} \, \|\mathbf{v} - \hat{\mathbf{v}}\|_2^2} \,. \tag{5.46}$$

Herein, the sample mean over the estimates of all Monte Carlo trials is used to approximate the expected value. The resulting average error is denoted by $\overline{\mathrm{RMSE}}(\mathbf{v}, \hat{\mathbf{v}})$.

### 5.7.3 Visualization of the PDL-OIAI/OMP Algorithms

The simulations in this subsection show the evolution of the PDL-OIAI and the PDL-OMP algorithms for different AE-iterations. If not stated differently, all simulations are carried out with an SNR of 10 dB. The dictionary has a coherence distance of $d_c = 14$. For CS, $M = 94$ Gaussian measurements (sampling matrix (I)) are taken, which reduces the data to be stored and processed by $(L-M)/L = 30\,\%$.
Figure 5.5(a) shows the simulated received sensor signal, $\mathbf{r}$ (noiseless), along with the estimates obtained by PDL-OIAI and PDL-OMP. Due to a small value of the effective receiver bandwidth (i.e. $\theta$), the reflections are broadened and overlap. Figure 5.5(b) depicts the original sparse signal, $\mathbf{x}$. The estimate, $\hat{\mathbf{x}}$, obtained by PDL-OIAI accurately identifies the original sparse support, $\mathcal{S}$. PDL-OMP cannot identify the correct support set due to strong IAI. Therefore, only $\hat{\mathcal{S}}$ obtained by PDL-OIAI can be used to estimate the reflection delays. Figure 5.5(c) illustrates the evolution of

the PDL-OIAI algorithm over several AE-iterations with indices $d$, $d = 1, \ldots, d^*$. It is zoomed on the closely-spaced sources on the left side and shows how the temporal sparse solutions approach the original sparse signal, $\mathbf{x}$. Finally, Figure 5.5(d) shows the temporal solution of the estimated dictionary parameter, $\hat{\theta}^{(d)}$, $d = 1, \ldots, d^*$, compared with the true value, $\theta = 1$. The linear decrease for $d < 5$ is due to a limited step size, which is introduced to improve the stability of the algorithms. After $d^* = 6$ iterations, PDL-OIAI estimates the dictionary parameter with high accuracy, while PDL-OMP converges to a stationary value with a small offset from the true value.

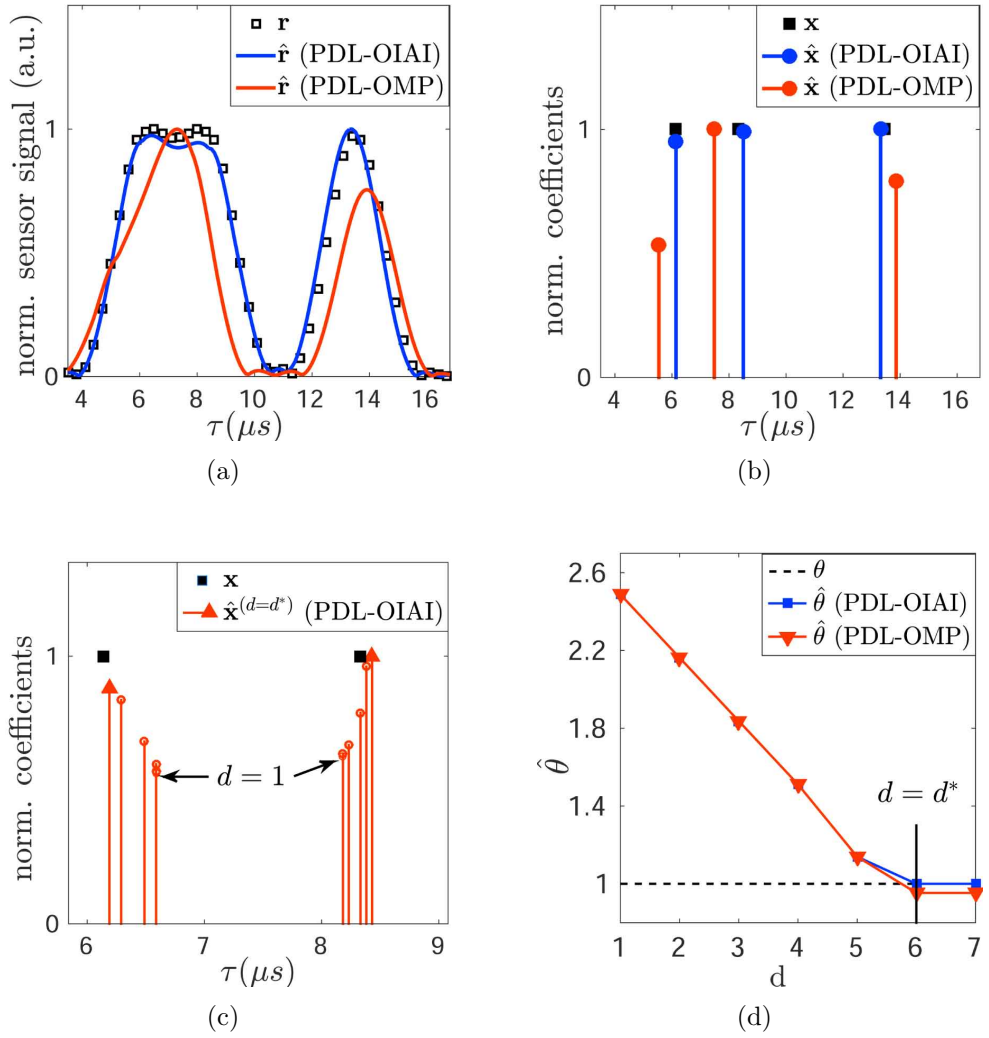



Figure 5.5. Visualization of the PDL-OIAI/OMP algorithms: (a) Simulated sensor signal (noiseless) and estimated sensor signal obtained by sparse reconstruction. (b) Original sparse signal (black bullets) and estimated sparse signal. (c) Temporal estimates of the sparse signal (zoomed on the closely-spaced reflections on the left side) for different AE-iterations. (d) Evolution of the dictionary parameter estimates for different AE iterations in comparison to th true value.

### 5.7.4 Performance Evaluation

The performance of the algorithms is evaluated in various scenarios of different SNRs, CS matrices, and coherence levels. The true sparse support is denoted by $\mathcal{S} = \text{supp}\{\mathbf{x}\}$ and $\hat{\mathcal{S}}$ is an estimate. The vector $\mathbf{s}$ and its estimate, $\hat{\mathbf{s}}$, are defined to contain the support indices in increasing order. The algorithms are initialized by $\hat{\theta}^{(0)} > \theta$, such that $d_c(\mathbf{D}(\hat{\theta}^{(0)})) < d_c(\mathbf{D}(\theta))$. Two values of the effective receiver bandwidth are considered. The coherence distance of the corresponding dictionaries is $d_c = 14$ and $d_c = 49$, respectively. For IAI mitigation, PDL-OIAI performs $J_U = 10$ internal iterations when $d_c = 14$, and to $J_U = 18$ when $d_c = 49$. All results are averaged over 500 Monte Carlo trials. In order to compare the estimation performance, the RMSE is calculated. The results are compared to the CRB in Section 5.4.3, which is calculated for a Gaussian CS matrix. The value of the CRB depends on the actual realization of the random matrix, $\mathbf{\Phi}$. Therefore, it is averaged over 1000 Monte Carlo trials.
It is to note that the CRB obtained this way is not the "true" CRB but the "averaged CRB", which is sometimes called the Miller and Chang bound (MCB) [156]. This bound only applies to locally unbiased estimators that are unbiased for all values of the random matrix (which is treated as a nuisance parameter). A detailed discussion can be found in [157].
The CRB also requires the derivative of $r(t,\theta)$ with respect to $\theta$. Since the dictionary is not a simple function of $\theta$, the derivative is approximated for a certain value, $\theta_0$. The $(l,i)$-th element in $\mathbf{D}'(\theta)$ is obtained by

$$\frac{\partial}{\partial\theta}[\mathbf{d}_i(\theta)]_l\bigg|_{\theta_0} \approx \frac{r(lT_d - i\delta t, \theta_0) - r(lT_d - i\delta t, \theta_0 - \Delta\theta)}{\Delta\theta}. \tag{5.47}$$

At first, the $\overline{\text{RMSE}}(\mathbf{s},\hat{\mathbf{s}})$ for the support indices is calculated, yielding the average number of grid points between the true and the estimated support indices. Next, the $\overline{\text{RMSE}}(\theta,\hat{\theta})$ is computed for the estimated dictionary parameter and compared to the CRB. Finally, the vectors $\mathbf{x}_\mathcal{S}$ and $\hat{\mathbf{x}}_\mathcal{S}$ are defined. They contain the coefficients in $\mathbf{x}$ and $\hat{\mathbf{x}}$, corresponding to the indices in $\mathcal{S}$ and $\hat{\mathcal{S}}$, respectively. Then, the $\overline{\text{RMSE}}(\mathbf{x}_\mathcal{S},\hat{\mathbf{x}}_\mathcal{S})$ is computed relative to the common amplitude, $A_x$, and compared with the CRB in Section 5.4.3. The comparison between the $\overline{\text{RMSE}}(\mathbf{x}_\mathcal{S},\hat{\mathbf{x}}_\mathcal{S})$ and the CRB is valid, since the constrained CRB equals the CRB of the 'oracle estimator', where of $\mathcal{S}$ is known.
In Figure 5.6, a Gaussian CS matrix is used and the performance is compared for the two dictionaries with coherence distances $d_c = 14$ and $d_c = 49$, respectively. According to Figure 5.6(a), PDL-OIAI estimates the reflection delays (i.e. $\mathcal{S}$) more accurately than PDL-OMP. It yields smaller values of $\overline{\text{RMSE}}(\mathbf{s},\hat{\mathbf{s}})$ for both dictionaries with $d_c = 14$ and $d_c = 49$, and for all SNRs. However, both algorithms are inaccurate when $d_c$ is too high. In Figure 5.6(b), the $\overline{\text{RMSE}}(\theta,\hat{\theta})$ of PDL-OIAI and PDL-OMP are similarly small for $d_c = 14$. They are both close to the CRB even at lower SNRs. For

$d_c = 49$, however, the dictionary coherence limits the performance of PDL-OMP. The performance gain due to IAI mitigation in PDL-OIAI is most significant at high SNRs. Figure 5.6(c) shows the $\overline{\mathrm{RMSE}}(\mathbf{x}_{\mathcal{S}}, \hat{\mathbf{x}}_{\hat{\mathcal{S}}})$, normalized by the common amplitude $A_x$. For $d_c = 14$, the performance of PDL-OIAI and PDL-OMP is comparable. For $d_c = 49$, PDL-OIAI is only slightly better than PDL-OMP.
Finally, in Figure 5.7 and Figure 5.8, the performance of the different CS matrices (I)-(III) is evaluated for $d_c = 14$, with $M/L = 50\%$ and $M/L = 20\%$, respectively. It is important to notice that the performance is similar for all CS matrices, although a (sparse) DF matrix discards 2/3 of all samples in every projection step. PDL-OIAI achieves the greatest improvement over PDL-OMP in estimating $\mathcal{S}$, yielding a significantly smaller value of the $\overline{\mathrm{RMSE}}(\mathbf{s}, \hat{\mathbf{s}})$. Regarding $\theta$ and $\mathbf{x}_{\mathcal{S}}$, an improvement over PDL-OMP can be observed only at higher SNRs. At lower SNRs, noise is the dominating factor. Generally, when the SNR is high, PDL-OIAI obtains accurate estimates of $\mathbf{s}$, $\theta$ and $\mathbf{x}_{\mathcal{S}}$, even though the number of CS measurements is small. For example, in Figure 5.8(a), the sample ratio is $M/L = 20\%$ and the $\overline{\mathrm{RMSE}}(\mathbf{s}, \hat{\mathbf{s}})$ is as small as 4 bins at SNR= 20 dB. Also, the $\overline{\mathrm{RMSE}}(\theta, \hat{\theta})$ in Figure 5.8(b) is close to the CRB. Nevertheless, it can be generally observed that $\overline{\mathrm{RMSE}}(\mathbf{x}_{\mathcal{S}}, \hat{\mathbf{x}}_{\hat{\mathcal{S}}})$ does not improve significantly at higher SNRs, leaving a constant gap to the CRB.

### 5.7.5 Experimental Validation Using Real Data

For an experimental validation, the performance of the CFS-DL framework is evaluated based on real data taken from the fiber sensing system in [70,86]. It was acquired at the Yamashita laboratory of photonic communication devices at the University of Tokyo, Japan. The core architecture follows the schematic in Figure 5.1. Further details of the experimental setup are given in [70, 86]. The parameters of the model in Section 5.5.2 are adjusted to match this reference system according to Section 5.7.1. The measured reflections are broadened due to a limited ADC bandwidth and due to an extended laser linewidth at high scan rates [70]. This broadening is taken into account by the effective receiver bandwidth, which is estimated in terms of the unknown dictionary parameter, $\theta = \Delta f$. In addition, there are a few distortions that are not explicitly modeled such as the signal-dependent noise amplitude and the skew shape of the reflections. The latter is caused by different rise and fall times in the temporal response of the PD.
The sensor signal, $\mathbf{r}$, consists of $L = 134$ samples of the measured sensor signal. CS is performed with a sampling ratio of $M/L = 40\%$. For the correct value of the dictionary parameter, the coherence distance is in the range $12 \leq d_c \leq 30$, while the considered range of dictionary parameters yields coherence distances $d_c \in [9, 89]$. The PDL-OIAI algorithm runs $J_U \leq 8$ iterations for IAI mitigation. The sensing fiber contains 4 FBGs, leading to $K = 4$ reflections with approximate delays of $[7.79, 9.05, 10.27, 12.30]$ $\mu$s.

Based on the delay grid of the dictionary, i.e. $\delta t = 50$ ns, the reflections are potentially off-grid. Each reflection is assumed to appear within a well-defined delay region.

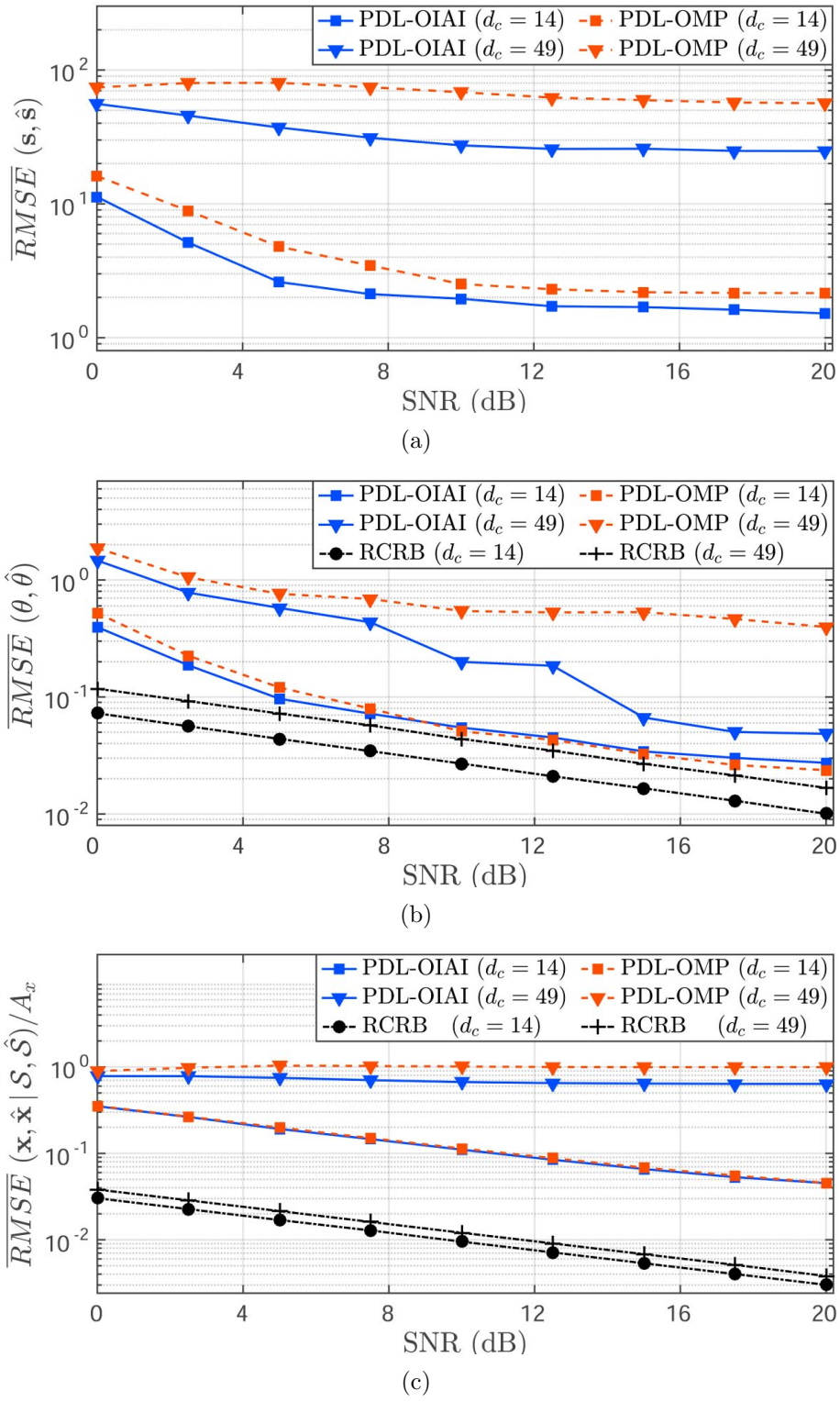
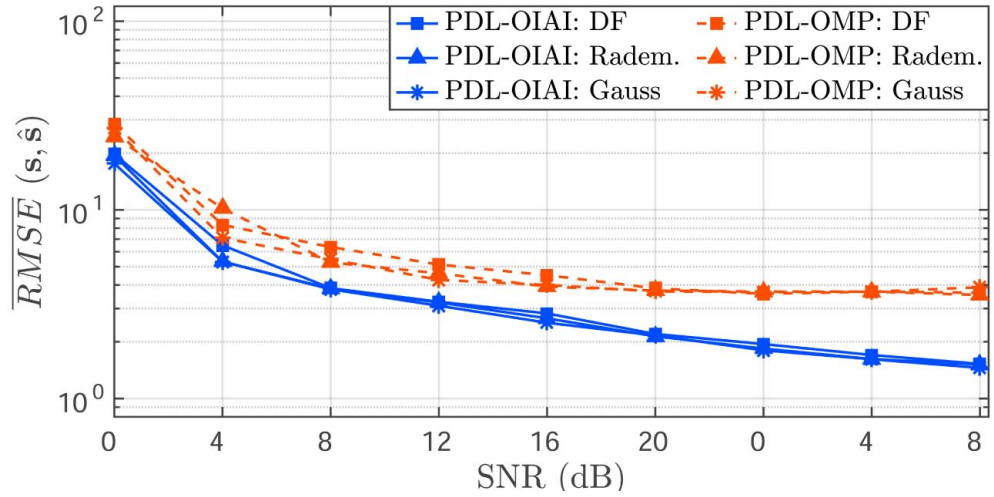
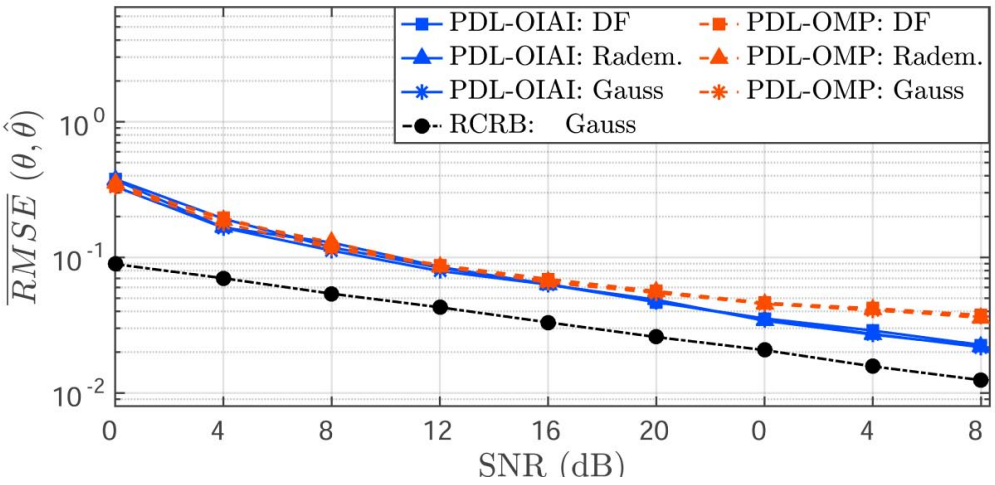


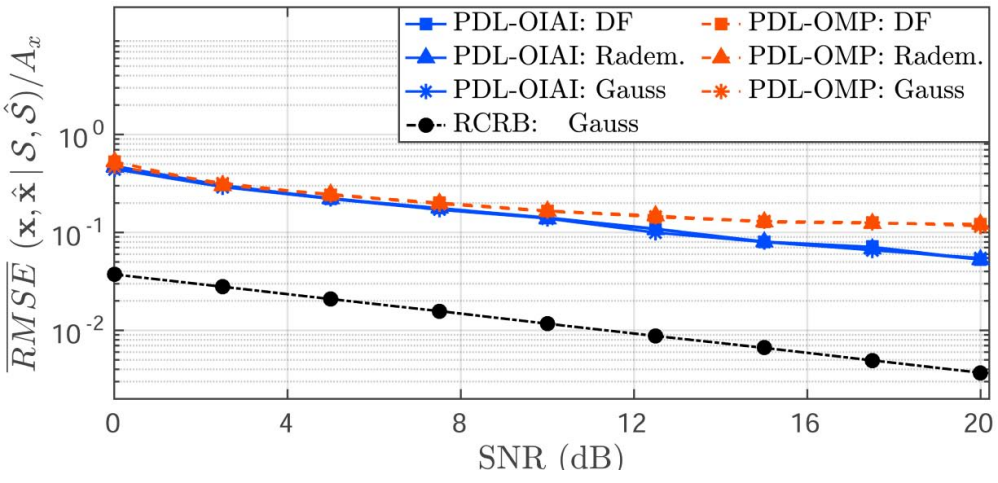

Figure 5.6. Performance of PDL-OIAI and PDL-OMP for different dictionaries with coherence levels $d_c = 14$ and $d_c = 49$, respectively.




(a)




(b)




(c)

Figure 5.7. Performance of the PDL-OIAI/OMP algorithms using 50% of the original samples and sampling matrices (I)-(III).

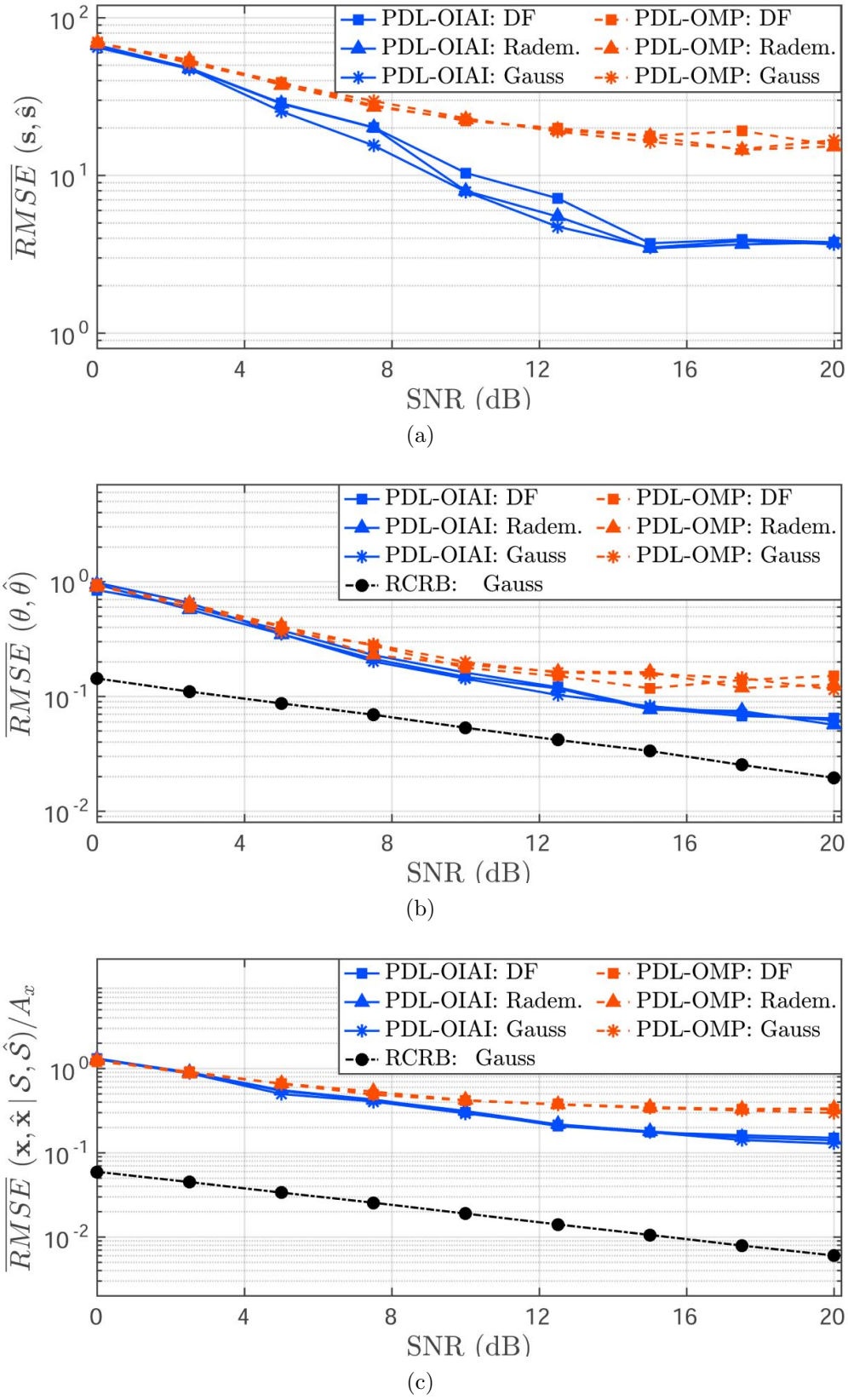



Figure 5.8. Performance of PDL-OIAI/OMP algorithms using 20% of the original samples and sampling matrices (I)-(III).

The randomness of the CS matrix and noise is taken into account by averaging the results of 500 Monte Carlo trials. Outliers may arise from numerical instabilities in the algorithms, e.g. due to a badly conditioned CS matrix. Therefore, the median of the estimates is shown along with the 85%-confidence intervals for $\mathcal{S}, \mathbf{x}_{\mathcal{S}}$, and for the estimated reflection signal.
Figure 5.9 shows the results obtained by PDL-OIAI and PDL-OMP. Each row corresponds to a different CS matrix (I)-(III). The left column of Figure 5.9 shows the measured sensor signal, $\mathbf{r}$. Also, the estimated reflection from the second FBG is shown in terms of the median over all Monte Carlo trials (solid and dashed line). The shaded areas depict the 85%-confidence interval. In the right column of Figure 5.9, the median of the estimates of the sparse coefficients, $\mathbf{x}$, is shown. The shaded areas signify the 85%-confidence intervals for the support, $\mathbf{s}$, while the vertical error bars signify the 85%-confidence intervals for the amplitudes of the significant coefficients, $\mathbf{x}_{\mathcal{S}}$.
Generally, PDL-OIAI and PDL-OMP show a comparable performance. This can be ascribed to a relatively high SNR and a small coherence distance of the correctly parametrized dictionary. However, the reflection delays, $\mathbf{s}$, and the reflection amplitudes, $\mathbf{x}_{\mathcal{S}}$, are slightly more accurate for PDL-OIAI. Both algorithms show some robustness to perturbations that are not explicitly modeled, as mentioned above. Yet, the confidence intervals imply a higher variance in the results obtained by PDL-OMP. Table 5.2 details the medians and the 85% confidence intervals for $\theta$ obtained by PDL-OIAI and PDL-OMP using the different CS matrices (I)-(III).

Table 5.2. Medians and 85%-confidence intervals for the effective receiver bandwidth, $\widehat{\Delta f}$ (in MHz), using the CS matrices (I)-(III). The ratio between the number of CS samples and the number of original samples is $M/L = 40\%$.

| Algorithm | Measure | Gaussian | Radem. | DF |
|---|---|---|---|---|
| **PDL-OIAI** | Median | 1.6560 | 1.6555 | 1.6458 |
| | $\Delta_{85\%}$ | 0.2565 | 0.2397 | 0.3727 |
| **PDL-OMP** | Median | 1.6545 | 1.6574 | 1.6371 |
| | $\Delta_{85\%}$ | 0.8139 | 0.8115 | 0.7745 |

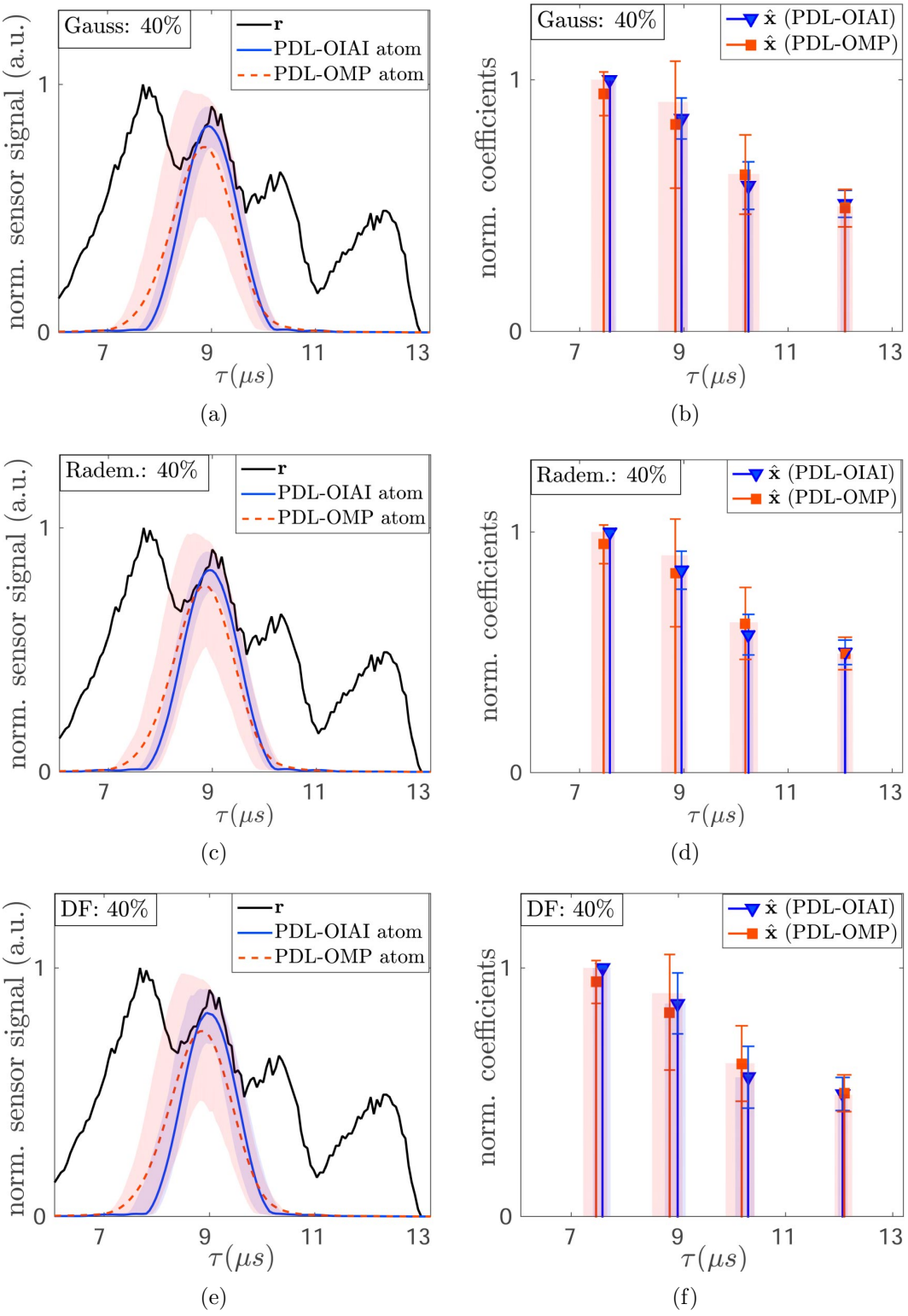



Figure 5.9. Experimental validation using real data. Left column: the received sensor signal (real data) and the median over all estimates of the second atom. The shaded areas denote the 85% confidence intervals. Right column: the median of the estimated support and amplitudes of $\hat{\mathbf{x}}$. The shaded areas and the vertical error bars indicate the 85% confidence intervals for estimating $\mathcal{S}$ and $\mathbf{x}_{\mathcal{S}}$, respectively. (a)-(b) Results obtained by a Gaussian CS matrix. (c)-(d) Rademacher CS matrix. (e)-(f) Database-friendly (DF) CS matrix.

## 5.8 Discussion and Findings

The CFS-DL framework is adopted for WDM-based quasi-distributed fiber-optic sensing. It is verified by experimental data that the considered signal model adequately describes the received signal of the system in [70, 86]. Therefore, the model is suitable to create a redundant translation-invariant dictionary for this application. In fact, the presented framework can be used for a wide range of applications, where the aim is to estimate the delay coefficients of localized signals, e.g. in laser ranging [11] or tomography [134]. A fundamental assumption of the considered model is the modular architecture, where the core system can be described by an LTI model. Based on this assumption, the model is generic and can be adapted to match different system setups as long as every component can be described by an LTI model.
Using CS, the number of samples to be stored and processed can be significantly reduced, according to the ratio $M/L$. In simulations and for experimental data, different CS matrices perform similarly well. A possible explaination for this observation is the high sparsity level of the signal with respect to the considered dictionary. Compared with Gaussian matrices, binary CS matrices such as Rademacher matrices offer a simpler hardware implementation. When the CS matrix is also sparse, a non-uniform sampling scheme can be adopted, such that the average sampling rate required to realize the CS projections is significantly reduced. In particular, DF matrices [106] lower the average sampling rate by $\approx 66\%$ and can be more efficiently stored than full sampling matrices. In addition, a number of $Q_A$ low-rate ADCs can be used in parallel. Then, a target sampling rate, $f_s$, can be realized when each ADC works at a constant rate, $f_s/Q_A$ [102]. As compared to other related AE-based sparse estimation and DL methods in [137,138,158], the additional IAI mitigation procedure in [32] enables PDL-OIAI to handle strong dictionary coherence. In each AE iteration, it allows for sparse estimation using a simple greedy OMP algorithm of low computational complexity, rather than using an optimization-based method. Moreover, a comparison to the reference method, PDL-OMP, emphasizes that IAI mitigation is essential in order to obtain reliable estimates of the dictionary parameter and for the sparse representation. It is shown that IAI mitigation yields significantly lower coherence distances, which eases the task of sparse estimation and, therefore, supports the work in [32]. In addition, the IAI mitigation process maintains the correspondence to the physical parameters of interest. That is, the sparse support (i.e. the reflection delays) estimated from the modified sensing dictionary, $\mathbf{W}$, can be used to estimate the FBG reflections with respect to the original dictionary, $\mathbf{D}$. In simulations and for experimental data, PDL-OIAI obtains reliable estimates of the sparse support, $\mathcal{S}$, the amplitudes of the sparse coefficients, $\mathbf{x}_{\mathcal{S}}$, and the dictionary parameter, $\theta$. The estimation performance for $\theta$ is even close to the CRB. Regarding the real data example, PDL-OIAI and PDL-OMP

show some robustness to model errors and work stably in the presence of certain non-modeled perturbations, e.g. the signal-dependent noise amplitude or the skew shape of the FBG reflections. Since OMP-based algorithms select only one atom per iteration, they are guaranteed to yield a $K$-sparse representation after $K$ iterations. Hence, PDL-OIAI and PDL-OMP are hardly affected when the underlying sparse signal is off-grid.

Similar to the works in [9, 24, 27], a composite parametric dictionary is considered. In CFS-DL, however, each sub-dictionary corresponds to the reflection from one FBG. Since only one reflection per FBG is expected, the composite dictionary yields a highly sparse representation of the signal, where the reflection delays are directly related to the sparse support, $\mathcal{S}$. Another general advantage of parametric shift-invariant dictionaries is that only the atom-generating function has to be stored. However, due to the IAI mitigation sub-routine in each AE iteration, also the full modified sensing dictionary, $\mathbf{W}$, needs to be stored.

A limitation of PDL-OIAI is the restriction to local optima in estimating the dictionary parameter. Also parameter identifiability can be a problem when different parameters are indistinguishable and cannot be individually estimated. However, it is sometimes useful to define auxiliary parameters that represent a common feature to which several indistinguishable parameters contribute. For example, the effective receiver bandwidth represents the broadening of the temporal reflections, which can be related to various parameters in the considered model. The high computational complexity of PDL-OIAI is mainly due to the IAI mitigation sub-routine. The modified sensing dictionary, $\mathbf{W}$, has to be computed in each AE iteration and also an initial regularization parameter is required. Nevertheless, it is shown that IAI mitigation is essential to handle strong dictionary coherence. In order to lower the computational complexity, the modified sensing dictionary, $\mathbf{W}$, can be calculated only once in the first OMP iteration. Moreover, the atoms of $\mathbf{W}$ can be calculated in parallel to achieve an additional speed-up.

In order to improve the estimation performance, some extensions of the PDL-OIAI can be considered. Higher SNRs can be achieved by jointly taking multiple CS measurement vectors into account and by applying the concept of group sparsity [159]. Sparsity over multiple measurements was previously used for sparse estimation in source localization problems [5]. It is also adopted for the direction finding problem in Chapter 4. Finally, the works in [5, 160] suggest to iteratively improve the grid accuracy using an initial coarse grid and, then, refining the grid in subsequent iterations.

## 5.9 Conclusion

An alternating sparse estimation and dictionary learning framework for highly coherent dictionaries is presented. The framework is adopted for the application of WDM-based quasi-distributed fiber-optic sensing with compressed sensing-based signal acquisition. A comparison between different CS matrices shows that the performance of a sparse *Database-friendly* (DF) sampling matrix is similar to that of a Gaussian sampling matrix. Sparse sampling matrices can be more efficiently stored and reduce the average sampling rate.

A parametric signal model is compiled to describe the sensor signal of a fiber sensor with a common core architecture. This model is used to generate the sparsity-promoting dictionary with translation-invariant structure. It is assumed that the system between the interrogating laser and the photodetector has a modular structure and that every component can be described by a linear time-invariant subsystem. Under this assumption, the model is generic and can be customized to describe a particular system setup. It is applicable to a wider class of applications besides fiber sensing.

Uncertainty in the dictionary is considered in terms of unknown global and local parameters. When different parameters are indistinguishable, an auxiliary parameter can be defined to describe the joint contribution of these parameters to a certain effect. In the context of fiber-optic sensing, the effective receiver bandwidth is defined as an auxiliary dictionary parameter that describes the broadening of the reflections in the temporal domain. The atoms of the dictionary are associated with the delays of the individual FBG reflections. Hence, the reflection delays can be determined by estimating the sparse support. A dense grid of delays is used to achieve high accuracy at the cost of a high dictionary coherence. When the number of non-zero entries in the sparse representation is small, conventional coherence measures, such as the mutual coherence [28] or the Babel function [9,12] (see Chapter 2.2.1.2), do not vary much for changes in the dictionary parameter or in the grid granularity of the dictionary atoms. Therefore, the coherence distance is used to emphasize the difference in the coherence level of the dictionary for different parametrizations.

The PDL-OIAI algorithm is introduced to accomplish alternating sparse estimation and dictionary learning in the presence of strong dictionary coherence. It is based on the *alternating estimation* (AE) paradigm and incorporates an *inter-atom-interference* (IAI) mitigation procedure to obtain a modified sensing dictionary with reduced coherence distance. In simulations and by experimental data, the performance of PDL-OIAI is evaluated in different scenarios. A comparison to a reference method without IAI mitigation, called PDL-OMP, shows that IAI mitigation is essential in the presence of strong dictionary coherence. As an AE-based algorithm, the dictionary parameters obtained by PDL-OIAI are locally optimal. Nevertheless, the estimation performance

for the considered dictionary parameter is close to the Cramér-Rao bound.
The problem of high computational complexity for IAI mitigation can be alleviated by parallel processing and by fixing the modified sensing dictionary after the first OMP iteration in each AE iteration.
Future work can be carried out to develop fast updating methods for the sparse solution, such that subsequent CS measurements can be efficiently incorporated. This enables the system to monitor non-stationary scenes such as time-varying perturbations. Also, the amplitudes of the sparse coefficients can be included in the estimation process. They contain information of the traveling distance based on the signal attenuation.

# Chapter 6

# Dictionary Learning Strategies Using a Probabilistic Sparse Model

In this chapter, hybrid and full Bayesian estimation strategies for sparse estimation and dictionary learning using highly redundant dictionaries are investigated. Due to high redundancy in the dictionary, it is assumed that the RIP conditions for basic $\ell_1$-minimization are not fulfilled.
In the preceding chapters, a deterministic model for the coefficients of the sparse representation is considered and the dictionary is modified in order to deal with redundancy. In this chapter, the structure of the dictionary is leveraged to design a probabilistic sparse model that allows for estimating a desired sparse representation from highly redundant dictionaries without modifying the dictionary itself. The sparse model developed in this chapter promotes selective shrinkage due to a suitably chosen prior density and achieves augmented sparsity by exploiting the structure of the dictionary. The resulting problem can be related to non-convex optimization with $\ell_p$-norm constraints, $0 < p < 1$, for which relaxed RIP requirements apply [161], such that the task of sparse estimation is alleviated albeit high dictionary redundancy.
Similar to the PDL-OIAI algorithm in Chapter 5, the presented hybrid Bayesian strategy follows the *alternating estimation* (AE) paradigm and employs a Monte Carlo Expectation Maximization algorithm to iteratively estimating the sparse representation and the dictionary parameters. The full Bayesian strategy, in contrast, is used to jointly estimate the dictionary parameters and the sparse representation.
The presented sparse estimation and dictionary learning methods are applied to the fiber sensing problem in Chapter 5. Moreover, the results are compared with those obtained by a deterministic sparse model using the PDL-OIAI algorithm in Chapter 5.

The material presented in this chapter is partly taken from [131][1].

### Chapter Outline

Section 6.1 gives an introduction and a motivation. Section 6.2 provides an overview of state-of-the-art methods and related work. In Section 6.3, the general problem is stated. In Section 6.4 the sparse model is developed and a relation to non-convex optimization is established. The complete hierarchical structure of the model and a hybrid Markov

[1] C. Weiss and A. M. Zoubir, "Dictionary Learning Strategies for Compressed Fiber Sensing Using a Probabilistic Sparse Model," submitted to *IEEE Transactions on Signal Processing, 2016*.

Chain Monte Carlo method for inference in this model are described in Section 6.5. The hybrid and full Bayesian dictionary learning strategies are introduced in Section 6.6. Section 6.7, evaluates the performance based on the problem of fiber sensing using simulations and real fiber sensing data. In order to compare the performance achieved by a probabilistic and a deterministic sparse model, the results are compared to those obtained by the PDL-OIAI algorithm in Chapter 5. Section 6.8 gives a discussion of the results and findings and Section 6.9 concludes this chapter.

## 6.1 Introduction and Motivation

There exist different approaches to estimating the parameters of a dictionary in a sparse estimation framework. Common statistical methods are *maximum likelihood* (ML) or *maximum a posteriori* (MAP) estimation [3]. The AE paradigm [53,54,162] (c.f. Chapter 5) is often considered for solving the resulting optimization problems. It is based on iteratively optimizing a local objective function and, therefore, yields locally optimal estimates. The *Expectation Maximization* (EM) algorithm [163] represents one variant of AE, which is often often employed to estimate unknown model parameters in hybrid Bayesian settings [3, 164, 165].

In sparse estimation, a deterministic or probabilistic model for the coefficients of the sparse representation can be considered. While the deterministic case is frequently encountered in the sparse estimation literature [9, 27], the flexibility of a probabilistic model can be advantageous, e.g. for the modeling of measurement errors and/or model errors. In addition, a hierarchical structure is useful when only vague prior knowledge is available. A comparative analysis between deterministic and probabilistic sparse models for a particular problem or application can support the selection process of a suitable model.

General advantages and disadvantages of hybrid and full Bayesian strategies are stated, e.g., in [164, 165]. Bayesian methods can yield better results if the introduced prior knowledge is sufficient and adequate [164]. Insufficient knowledge can lead to a small-sample bias [164]. Also, the computational complexity can be high.

In the problem of fiber sensing, only few samples are acquired due to the CS-based acquisition scheme. While no prior knowledge of $\theta$ is available, strong sparsity-related prior knowledge is incorporated.

Different probabilistic models have been proposed to promote sparsity. In principle, weakly sparse and strongly sparse models (discrete mixtures) can be distinguished [162]. One important class of strongly sparse models is the family of 'Spike & Slab' models [166]. However, practical signals are typically corrupted by measurement errors or

noise, such that the underlying signal is not exactly sparse. When these effects are neglected, they can affect the quality of the sparse representation by introducing a bias. This effect can be observed when sparse regularization is used to suppress the contributions of measurement errors or noise in the obtained sparse representation (c.f. Chapter 4). Therefore, a weakly sparse model is considered in this chapter. Weakly sparse models are based on continuous distributions that are peaked around zero with heavy tails such as Gaussian or double-exponential (Laplace) distributions [162]. They favor small values but do not prevent selected elements from becoming sufficiently large [162, 167]. Nonetheless, when the dictionary is highly coherent, the selective shrinkage induced by the sparsity prior alone might not yield a sufficiently sparse representation. Additional knowledge of the data or the dictionary structure, such as correlation, can help to obtain a desired sparse representation [168, 169].

In complicated probabilistic models, it can be unfeasible to evaluate the posterior distribution. Yet, approximate inference is possible using *Markov Chain Monte Carlo* (MCMC) or variational Bayes methods [163, 170, 171]. Variational methods are advantageous in terms of scalability and computational complexity. They attempt to approximate the posterior distribution by simple analytic functions, thereby introducing a deterministic approximation error [163, 171]. MCMC methods generate samples of the posterior distribution, where the samples form a Markov chain [163]. Although MCMC can be costly when the sample space is high-dimensional, there do exist efficient sampling algorithms such as the *Hamilton Monte Carlo* (HMC) method. The quality of an approximation obtained by sampling methods increases with the allowed run-time, which leads to a stochastic approximation error [163].

One of the aims in this chapter is to provide a comparative analysis between the sparse representations obtained by a deterministic and a probabilistic sparse model. Therefore, the developed estimation methods in this chapter are applied to the problem of compressed fiber sensing in Chapter 5, using the corresponding translation-invariant dictionary in Equation 5.1. The signal model in Chapter 5 is modified by introducing a probabilistic hierarchical (weakly sparse) model for the coefficients of the sparse representation. It is shown that the considered sparse model can be related to non-convex optimization with $\ell_p$-norm constraints, where $0<p<1$. In addition, the structure of the data and the dictionary is leveraged to achieve augmented sparsity. In particular, a Markov Random Field model is used to describe the dependency between the coefficients of the sparse representation, which induces additional collective shrinkage effects. According to the fiber sensing problem in Chapter 5, an unknown global parameter is used to describe uncertainty in the dictionary. Two strategies are presented to estimate this parameter. The first strategy, referred to as ***S1***, considers a deterministic parameter model and employs a Monte Carlo EM algorithm to estimate the dictionary parameter. The second strategy is referred to as ***S2*** and considers a probabilistic hier-

archical model for the dictionary parameter. This leads to a fully probabilistic model, where the sparse representation and the dictionary parameter are jointly estimated. Since the dimensionality of the sparse representation can be high, HMC and Gibbs sampling are used to accomplish approximate inference. Using simulations and experimental data, the results are compared to those obtained based on the deterministic sparse model in Chapter 5. The *Cramér-Rao bound* (CRB) derived in Chapter 5 serves as a benchmark to assess whether or not a performance gain is achieved by using the probabilistic sparse model presented in this chapter.

## 6.2 State of the Art and Related Work

Different probabilistic sparsity-promoting models have been discussed in the literature. A comprehensive overview and an analysis of different sparse prior models with hierarchical structure are provided in [172]. Hierarchical sparse models have been used in various applications such as direction-of-arrival estimation [173] or *compressed sensing* (CS) [174,175]. Among the sparsity-promoting models, weakly sparse and strongly sparse models can be distinguished. The 'Spike & Slab' model is a prominent variant of a strongly sparse model, which has been investigated, e.g., in [166, 176, 177]. Also weakly sparse models have been widely studied, e.g., in [167, 171]. In [174, 175], a Bayesian framework for CS with Laplace prior is proposed, and the work in [178] focuses on CS with unknown noise level and Gaussian sparsity prior. The authors in [167] address the general problem of selecting an appropriate prior for sparse regression and shrinkage, including an overview of the properties of different priors. They point out that the problem can be viewed in terms of choosing a prior distribution or a penalty function. An analysis of various Bayesian models and a comparison to $\ell_1$-minimization is provided in [162]. Seeger [171] observed favorable shrinkage properties of certain priors, e.g. the Laplace prior, where small components are strongly favored but selected components happened to be significantly larger. The authors in [177] call this property 'selective shrinkage'. It can be generally observed for heavy-tailed priors such as the 'horseshoe' prior in [167,179] or the Student's $t$-prior in [171]. It is shown in [180] that, compared to a Gaussian or Laplace prior, the horseshoe prior shrinks significant sparse coefficients less and small coefficients stronger to zero. These findings build the basis for the sparsity model in this chapter, where the shape of the considered sparsity prior is close to that of the horseshoe prior.

Instead of selecting sparsity priors, other works focus on the perspective of penalized regression. In [181], higher sparsity levels have been reported by imposing a penalty on the $\ell_p$-norm for $0 < p < 1$, rather than penalizing the $\ell_1$-norm. It has been shown

in [161] that the RIP requirements can be relaxed in the case of $\ell_p$-norm penalization. Also, the authors in [161, 182] found that non-convex CS based on $\ell_p$-norm penalization requires less measurements than standard CS based on the $\ell_1$-norm. Inspired by these results, a relationship between the sparse model presented in this chapter and non-convex optimization with $\ell_p$-norm constraints is established.
In addition to selecting a sparsity prior, the structure of the signal and the dictionary can be leveraged. The concept of 'block sparsity' [168] is a popular example. It has been applied for Bayesian learning, e.g. in [169]. A related concept of 'joint sparsity' is used by Wakin *et al.* in [183]. The correlation between the underlying sparse signal of subsequent CS measurements is exploited in [5], with the aim to improving the signal-to-noise ratio. Correlation structures can also be used to achieve smoothness [184] or to describe the relation between adjacent pixels in image processing [185]. Other works consider sparsity directly in the correlation domain [186, 187]. In this chapter, correlation between the sparse coefficients is modeled in terms of Markov Random Fields [188], which yields additional collective shrinkage effects.
The work in [162] emphasizes that MCMC methods are powerful for inference in sparse models. The *Hamilton Monte Carlo* (HMC) method [189] is shown to be efficient in high dimensional spaces and in the presence of correlation. A combination of HMC and Gibbs sampling is suggested by the author in [189]. This technique is applied for inference in the hierarchical sparse model presented in this chapter. In order to estimate dictionary parameters along with a sparse representation, AE is frequently employed [138, 158]. The presented dictionary learning strategy, ***S1***, also represents one AE variant. Different from AE-based estimation, a full Bayesian framework for sparse estimation and dictionary learning is proposed in [56]. This framework can be related to the presented method, ***S2***, but it does not consider correlation among the sparse coefficients.

## 6.3 Problem Statement

The problem of estimating the reflection coefficient of localized signals is addressed. It is cast as a sparse estimation and dictionary learning problem according to Chapter 5, Section 5.3. A parametric translation-invariant dictionary, $\mathbf{D}(\theta)$, with high redundancy and an unknown scalar parameter, $\theta$, is considered. It is assumed that $\theta$ describes the localization of the modeled signals, such that it impacts the the width of the region to which the modeled signals are confined. The $i$-th dictionary atom is created from shifted versions of a generating function, $r(t, \theta)$, which is sampled at a design sampling rate, $T_d$, i.e.

$$[\mathbf{d}_i]_l(\theta) \;=\; r(lT_d - i\,\delta t, \theta), \qquad l = 1, \ldots, L, \qquad i = 1, \ldots, N, \tag{6.1}$$

where $\delta t$ denotes the spacing between subsequent translation coefficients. It determines the granularity of the dictionary grid, hence, the redundancy of the dictionary. Note that the relation between $\theta$ and $\mathbf{D}$ is the same as described in detail in Chapter 5.5.2. CS is used to acquire the observed signal. The sampling process is described by a CS sampling matrix, $\mathbf{\Phi}$, and the signal model for the CS measurements, $y_m,\ m = 1, \ldots, M$, is given by

$$\mathbf{y} \;=\; \mathbf{\Phi D}(\theta)\mathbf{x} + \mathbf{n} \;=\; \mathbf{B}(\theta)\mathbf{x} + \mathbf{n}\,. \tag{6.2}$$

Herein, $\mathbf{B}(\theta)$ denotes the combined sensing dictionary and $\mathbf{x} \in \mathbb{R}^N$ is a sparse vector with $K$ significant components at the index positions in $\mathcal{S} = \{i_1, \ldots, i_K\}$. The vector $\mathbf{n} \in \mathbb{R}^M$ represents a Gaussian noise component with independent and identically distributed entries, $n_m \sim \mathcal{N}(0, \sigma_n^2),\ m = 1, \ldots, M$. The likelihood function can be written by

$$p(\mathbf{y}\,|\,\mathbf{x}, \theta) \;=\; (\sqrt{2\pi}\sigma_n)^{-M} \exp\!\left(-\frac{1}{2\sigma_n^2}\|\mathbf{y} - \mathbf{\Phi D}(\theta)\mathbf{x}\|_2^2\right). \tag{6.3}$$

In the context of the considered problem in fiber-optic sensing (c.f. Chapter 5), $r(t, \theta)$ describes the portion of the interrogating signal that is reflected from a *fiber Bragg grating* (FBG). The parameter $\theta \in \mathbb{R}_+$ describes the effective receiver bandwidth. It is defined as an auxiliary parameter to describe the temporal broadening of the received reflections. For a fixed spacing, $\delta t$, the parameter $\theta$ directly impacts the similarity between adjacent dictionary elements. A small value of $\theta$ corresponds to broad reflections and high similarity.
The overall task is to estimate the dictionary parameter, $\theta$, along with the underlying sparse representation, $\mathbf{x}$, that correctly describes the FBG reflections.

## 6.4 Probabilistic Sparse Model for Localized Signals

In order to correctly estimate the translation coefficients of localized signals, the localization, i.e. the width of the region to which the signals are confined, has to be correctly described by the generating function of the translation-invariant dictionary. Without prior assumptions, there exist indeed many valid explanations of the observed data based on the general model in (6.2), and the dependency of the sparse representation, $\mathbf{x}$, on the localization of the signal, $\theta$, (and vice versa) introduces additional ambiguity.
This dependency can be illustrated for the problem of fiber sensing: a large value of the dictionary parameter, $\theta$, corresponds to a large receiver bandwidth, such that the

dictionary models highly confined and sharp reflections peaks. However, when the observed data is generated by a number of $K$ physical reflections that are significantly broader, then $\tilde{K} \gg K$ non-zero coefficients in $\mathbf{x}$ are necessary to fit the data based on this dictionary. In this case, the interpretation with respect to the considered model is misleading, since the individual peaks are no longer related to the physical reflections. Therefore, a suitable parametrization of the dictionary requires a smaller value of $\theta$, yielding a sparse representation with exactly $K$ significant non-zero coefficients. In this setting, the delays of the FBG reflections (i.e. the translation coefficients of the observed localized signals) can be correctly estimated from the positions of the non-zero coefficients in $\mathbf{x}$, i.e. from the elements in $\mathcal{S}$. This example shows that, when both $\mathbf{x}$ and $\theta$ are unknown, the translation coefficients can only be estimated when the sparsity assumption is properly incorporated in terms of a suitable prior.

Measurement errors, noise or other distortions can cause artifacts in the sparse representation. A 'strongly sparse' model (discrete mixture) enforces a truly sparse representation by completely ignoring these effects. This can lead to a biased representation that does not correctly indicate the translation coefficients (i.e. the reflection delays in the fiber sensing problem). A 'weakly sparse' model, in turn, yields a representation in which most coefficients are close to but not exactly zero. Then, the translation coefficients more accurately indicated by the most significant components, since their positions are less affected by distortions. Nevertheless, high redundancy and coherence in the dictionary complicates the task of sparse estimation based on a weakly sparse model. In a translation-invariant dictionary, directly adjacent elements exhibit the highest similarity. Therefore, bursts of larger coefficients in $\mathbf{x}$ appear around the true translation coefficients if the prior does not induce sufficient selective shrinkage. This is indeed the case for commonly used priors, e.g. Laplace priors, leading to ambiguous representations.

A possible explanation of this observation can be found by considering the dual interpretation of classical $\ell_1$-minimization and *maximum a posteriori* (MAP) estimation. In order to estimate a sparse representation in (6.2) using $\ell_1$-minimization, the corresponding optimization problem has to be solved. However, for a likelihood function as in (6.3), the same optimization problem has to be solved to determine the MAP solution of $\mathbf{x}$, given that the entries in $\mathbf{x}$ are independent and each element is assigned a Laplace prior [162]. In the scope of this analogy, it can be argued that the ambiguity in the sparse representation is due to the fact that the RIP requirements for $\ell_1$-minimization are not fulfilled for highly coherent dictionaries [13, 62]. It has been shown in [161, 182], that the RIP requirements for $\ell_p$-minimization can be relaxed if $0 < p < 1$, which results in a non-convex optimization problem. This result suggests that a sparsity prior for which the MAP problem can be related to non-convex $\ell_p$-minimization is more suitable to deal with high dictionary coherence.

The choice of a suitable prior may also depend on application-specific restrictions. In the problem of fiber sensing, the observed reflections are proportional to the optical power. Since the considered dictionary models these reflections, the coefficients of the sparse representation are constrained to be non-negative. Based on these requirements, an appropriately parametrized Weibull prior that resembles a positive version of the horseshoe prior in [167] is considered $\forall\ i = 1, \ldots, N$:

$$x_i \ \sim \ p(x_i) \ = \ \mathcal{W}(x_i \,|\, \lambda_w, k_w) \ = \ \left(\frac{k_w}{\lambda_w^{k_w}}\right) x_i^{k_w - 1} \exp\left(-\left(\frac{x_i}{\lambda_w}\right)^{k_w}\right), \quad x_i \geq 0\,, \quad (6.4)$$

where $\lambda_w, k_w$ are the scale and shape parameters, respectively. The joint prior density of $\mathbf{x}$ is given by

$$p(\mathbf{x} \,|\, k_w, \lambda_w) \ = \ \prod_{i=1}^{N} \mathcal{W}(x_i \,|\, \lambda_w, k_w) \ = \ \left(\frac{k_w}{\lambda_w^{k_w}}\right)^N \left(\prod_{i=1}^{N} x_i^{k_w - 1}\right) \exp\left(-\lambda_w^{-k_w} \sum_{i=1}^{N} x_i^{k_w}\right) \quad (6.5)$$

with $\mathbf{x} \succeq \mathbf{0}$. A qualitative sketch of this prior in the bivariate case is shown in Fig. 6.1(a).

On the basis of (6.3) and (6.5), the problem of sparse estimation can be related to a constrained ML estimation problem. This can be seen by first formulating the MAP estimation problem, i.e.

$$\begin{aligned} \hat{\mathbf{x}}_{\text{MAP}} \ &= \ \arg\max_{\mathbf{x} \succeq \mathbf{0}} \ \log(\, p(\mathbf{y} \,|\, \mathbf{x}, \theta)\, p(\mathbf{x} \,|\, k_w, \lambda_w)) \\ &= \ \arg\min_{\mathbf{x} \succeq \mathbf{0}} \ -\log(\, p(\mathbf{y} \,|\, \mathbf{x}, \theta)\, p(\mathbf{x} \,|\, k_w, \lambda_w)) \\ &\propto \ \arg\min_{\mathbf{x} \succeq \mathbf{0}} \ \|\mathbf{y} - \mathbf{\Phi}\mathbf{D}(\theta)\mathbf{x}\|_2^2 \ + \ \breve{\mu}_1 \sum_{i=1}^{N} \log(x_i) \ + \ \breve{\mu}_2 \sum_{i=1}^{N} x_i^{k_w}\,, \end{aligned} \quad (6.6)$$

where $\breve{\mu}_1 = (1 - k_w)$, $\breve{\mu}_2 = \lambda_w^{-k_w}$ with $0 < k_w < 1$, and $\breve{\mu}_1, \breve{\mu}_2 > 0$. The coefficients $\breve{\mu}_1, \breve{\mu}_2$ can be interpreted as regularization parameters.

Using (6.6), a related constrained ML problem can be formulated by defining two functions,

$$g_1 \ = \ \sum_{i=1}^{N} x_i^{k_w} \ - \ \lambda_1^{k_w} \qquad \text{and} \qquad g_2 \ = \ \sum_{i=1}^{N} \log(x_i) \ - \ \lambda_2\,. \quad (6.7)$$

These functions can be used to define inequality constraints of the form $g_1 \leq 0$ and $g_2 \leq 0$, where $\lambda_1, \lambda_2 \in \mathbb{R}_+$ are hyperparameters that can be related to the regularization parameters $\breve{\mu}_1$ and $\breve{\mu}_2$, respectively. Their purpose is to restrict the search space of the solution in order to account for the shrinkage effects of the considered prior in the unconstrained formulation in (6.6).

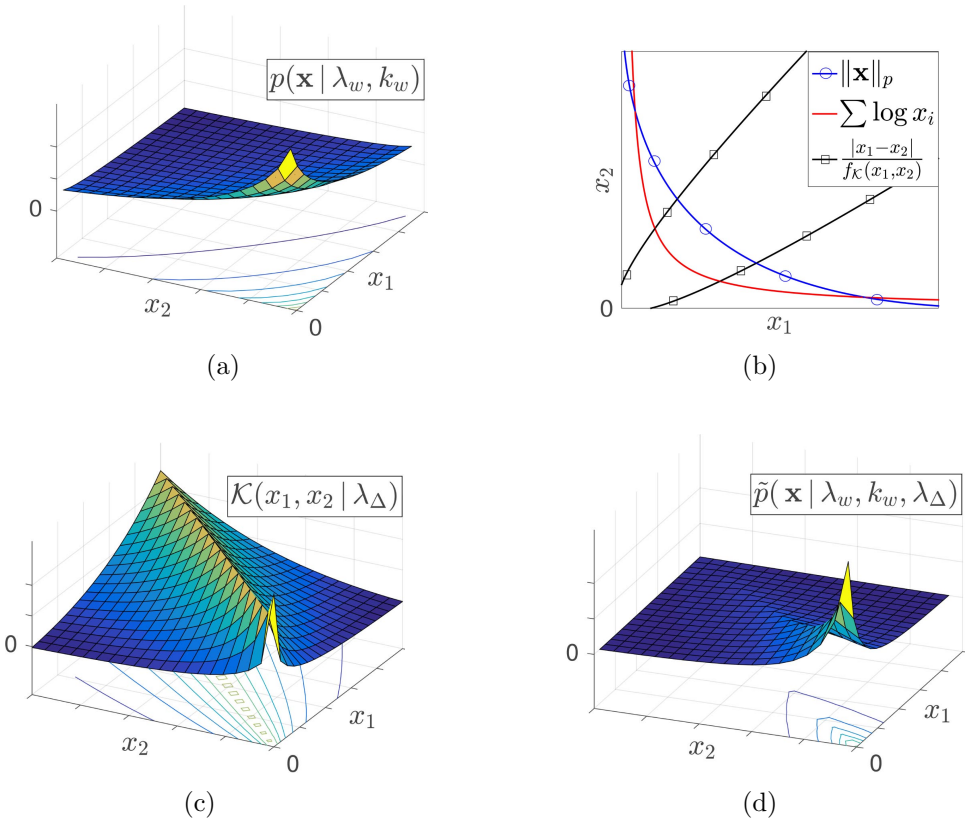



Figure 6.1. Bivariate examples. (a) Weibull prior $p(\mathbf{x} \mid \lambda_w, k_w)$, (b) search space of the constrained non-convex optimization problem, (c) local similarity function $\mathcal{K}(x_1, x_2 | \lambda_\Delta)$, (d) modified joint density $\tilde{p}(\mathbf{x} | k_w, \lambda_w, \lambda_\Delta)$ with $\mathbf{x} = [x_1, x_2]^\top$ and $f_\mathcal{K}(x_1, x_2) = \sqrt{x_1 + x_2}$.

On this basis, a related constrained version of the ML problem can be formulated by

$$\arg\min_{\mathbf{x} \succ \mathbf{0}} \quad \|\mathbf{y} - \mathbf{\Phi}\mathbf{A}(\theta)\mathbf{x}\|_2^2 \tag{6.8}$$

$$\text{s.t.} \quad \|\mathbf{x}\|_{k_w} \leq \lambda_1 \tag{6.9}$$

$$\text{and} \quad \sum_{i=1}^N \log(x_i) \leq \lambda_2\,, \tag{6.10}$$

where $\|\mathbf{x}\|_p = (\sum_{i=1}^N |x_i|^p)^{1/p}$ denotes the $\ell_p$-norm with $p = k_w < 1$. This problem is non-convex due to the non-convex constraint set and the sparsity of the solution is controlled by the hyperparameters, $\lambda_1$ and $\lambda_2$. The search space in the bivariate case, restricted by the constraints (6.9)-(6.10), is depicted in Fig. 6.1(b) for fixed values of $\lambda_1$ and $\lambda_2$.

### 6.4.1 Local Covariance Model for Augmented Sparsity

The problem of high dictionary coherence is addressed by exploiting the specific structure of the shift-invariant dictionary and the signal. Inspired by the concept of block sparsity [168], the aim is to promote sparsity among groups of variables, such that the positions of the translation coefficients of the signals can be correctly identified. It is assumed that the $K$ non-zero coefficients corresponding to the observed reflections are well-separated. Consequently, the significant components in $\mathbf{x}$ are surrounded by larger groups of non-significant coefficients. It is further assumed that all non-significant coefficients are similarly small as compared to the significant components. This local similarity property between adjacent elements can be described in terms of a Markov Random Field model [188]. To this end, a prior can be assigned to the differential coefficients, $\Delta x_i = x_{i+1} - x_i$, $i = 1, \ldots, N-1$, in order to suppress large changes in the amplitudes of adjacent coefficients in $\mathbf{x}$. Since most non-significant coefficients are small, occasionally appearing larger amplitudes among the non-significant coefficients are equalized. This effect is referred to as 'collective shrinkage'. However, without further modifications, also significant coefficients are suppressed in this model. Therefore, a restriction for the change in the amplitudes has to be locally specified by incorporating the amplitudes themselves.
With this in mind, a kernel function can be specified for all adjacent pairs of coefficients in $\mathbf{x}$, i.e. for all $x_i, x_{i+1},\ i = 1, \ldots, N-1$:

$$\mathcal{K}(x_i, x_{i+1} \,|\, \lambda_\Delta) \;=\; \exp\!\left(-\lambda_\Delta \frac{|x_{i+1} - x_i|}{f_\mathcal{K}(x_i, x_{i+1})}\right), \tag{6.11}$$

The bivariate functions,$f_\mathcal{K}$, depend directly on the amplitudes of the sparse coefficients. Together with $\lambda_\Delta$, these functions control the allowed amplitude variation between adjacent coefficients. Subsequently, it is assumed that $f_\mathcal{K}(x_i, x_{i+1}) = (x_i + x_{i+1})^r / N_x$, $i = 1, \ldots, N-1$, where $r \leq 1$ and $N_x < \infty$ are positive constants.
When this model is incorporated, the modified joint prior density of the sparse coefficients becomes

$$\tilde{p}(\mathbf{x} \,|\, k_w, \lambda_w, \lambda_\Delta) \;=\; \frac{1}{Z_\mathcal{K}} \mathcal{W}(x_N \,|\, k_w, \lambda_w) \times \prod_{i=1}^{N-1} \mathcal{K}(x_i, x_{i+1} \,|\, \lambda_\Delta)\, \mathcal{W}(x_i \,|\, k_w, \lambda_w), \tag{6.12}$$

where $Z_\mathcal{K}$ is a normalization constant. To verify that $\tilde{p}(\mathbf{x} \,|\, k_w, \lambda_w, \lambda_\Delta)$ can be normalized, consider some positive variables $\alpha, \beta \in \mathbb{R}_+$. Then, it follows from (6.11) and from the definition of $f_\mathcal{K}(x_i, x_{i+1})$ that $\mathcal{K}(x_i, x_{i+1} \,|\, \lambda_\Delta)$ is bounded, i.e.

$$0 \;<\; \mathcal{K}(\alpha, \beta \,|\, \lambda_\Delta) \;=\; \mathcal{K}(\beta, \alpha \,|\, \lambda_\Delta) \;\leq\; 1\,. \tag{6.13}$$

Hence, there exists a positive constant $Z_\mathcal{K} < \infty$ for which $\tilde{p}(\mathbf{x} \,|\, k_w, \lambda_w, \lambda_\Delta)$ can be normalized. In Figures 6.1(c)-6.1(d), the kernel function, $\mathcal{K}(x_i, x_{i+1} \,|\, \lambda_\Delta)$, and the modified joint prior density of $\mathbf{x}$ (bivariate case) are depicted , respectively.

A relation between the modified prior density of $\mathbf{x}$ and the constrained optimization problem in (6.8)-(6.10) can be established by considering an additional constraint:

$$|x_{i+1} - x_i| \;\leq\; \mu_i, \qquad i = 1, \ldots, N-1\,, \tag{6.14}$$

where $\mu_i$ depends on $f_{\mathcal{K}}(x_i, x_{i+1})$ and $\lambda_\Delta$. A bivariate example of the restricted search space for fixed values of $\mu_1$ is depicted in Figure 6.1(b).
The relation of this sparse model to Markov Random Fields can be illustrated in terms of the conditional densities for the coefficients $x_i$, $i = 1, \ldots, N$. In order to simplify notations, let $p(x_i \,|\, \mathbf{x}_{\setminus i}) = p(x_i \,|\, x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_N)$. Then, the conditional prior densities are given by

$$\begin{aligned}
\tilde{p}(x_i \,|\, \mathbf{x}_{\setminus i}, k_w, \lambda_w, \lambda_\Delta) &= \tilde{p}(x_i \,|\, x_{i-1}, x_{i+1}, k_w, \lambda_w, \lambda_\Delta) \\
&\propto \mathcal{W}(x_i \,|\, k_w, \lambda_w)\, \mathcal{K}(x_{i-1}, x_i \,|\, \lambda_\Delta)\, \mathcal{K}(x_i, x_{i+1} \,|\, \lambda_\Delta)\,, \qquad (6.15) \\
\tilde{p}(x_1 \,|\, \mathbf{x}_{\setminus 1}, k_w, \lambda_w, \lambda_\Delta) &\propto \mathcal{W}(x_1 \,|\, k_w, \lambda_w)\, \mathcal{K}(x_1, x_2 \,|\, \lambda_\Delta), \qquad (6.16) \\
\tilde{p}(x_N \,|\, \mathbf{x}_{\setminus N}, k_w, \lambda_w, \lambda_\Delta) &\propto \mathcal{W}(x_N \,|\, k_w, \lambda_w)\, \mathcal{K}(x_{N-1}, x_N \,|\, \lambda_\Delta)\,. \qquad (6.17)
\end{aligned}$$

It can be seen that dependencies appear only between directly adjacent coefficients.

In order to complete the sparse model, a hierarchical structure is considered to account for uncertainty in prior assumptions. To this end, (conjugate) inverse Gamma (Inv-$\Gamma$) priors are assigned to the randomized scale parameters, $\lambda_w$ and $\lambda_\Delta$. Given $\lambda_w$, the prior of the shape parameter, $k_w > 0$, is chosen as in [190], i.e.

$$p(k_w \,|\, a', b', (d')^{k_w}, \lambda_w) \;=\; \frac{k_w^{a'}}{Z_{k_w}} \exp\!\left(-b' k_w - \frac{(d')^{k_w}}{\lambda_w}\right), \tag{6.18}$$

where $Z_{k_w}$ is a normalization constant. Figure 6.2 shows a factor graph that visualizes the complete structure of the sparse model.

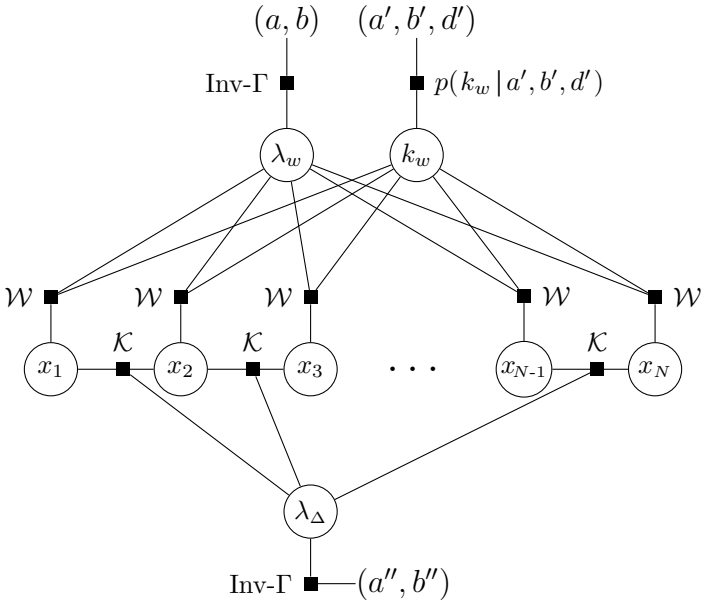



Figure 6.2. Factor graph of the model for the sparse coefficients with local similarity kernels.

## 6.5 Hybrid Markov Chain Monte Carlo Algorithm

Inference in the proposed model is accomplished using a hybrid MCMC technique that combines HMC and Gibbs sampling. The advantages of HMC are (i) it requires only an analytic expression for the posterior density to be sampled and (ii) it is efficient for sampling high-dimensional spaces in the presence of correlation [189]. Nevertheless, it is stated in [189] that the performance is limited when hyperparameters are included in the sampling process. They often have a sharply peaked posterior distribution, which requires small step sizes in the HMC algorithm. Therefore, following the suggestions in [189], Gibbs sampling is used to sample the hyperparameters separately. In addition, the noise variance, $\sigma_n^2$, is inferred. It is assigned an inverse Gamma (Inv-$\Gamma$) prior and included in the Gibbs sampling procedure. The joint density of the sparse coefficients, $\tilde{p}(\mathbf{x} \,|\, \lambda_w, k_w, \lambda_\Delta)$, is sampled using HMC, keeping all other variables fixed. The overall model is stated below:

$$\begin{aligned}
\mathbf{x} \,|\, k_w, \lambda_w, \lambda_\Delta &\sim \tilde{p}(\mathbf{x} \,|\, k_w, \lambda_w, \lambda_\Delta) && \text{in (6.12)},\\
\lambda_w &\sim \text{Inv-}\Gamma(\lambda_w \,|\, a, b),\\
k_w \,|\, \lambda_w &\sim p(k_w \,|\, a', b', (d')^{k_w}, \lambda_w) && \text{in (6.18)},\\
\lambda_\Delta &\sim \text{Inv-}\Gamma(\lambda_\Delta \,|\, a'', b'')\\
\sigma_n^2 &\sim \text{Inv-}\Gamma(\lambda_\Delta \,|\, a_\sigma, b_\sigma)\,. && (6.19)
\end{aligned}$$

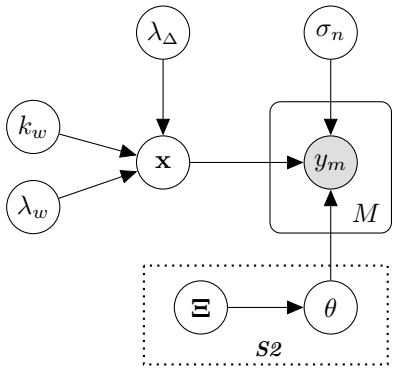



Figure 6.3. Graphical representation of the latent variable model.

Herein, it is assumed that the variables $\mathbf{x}, \sigma_n^2$ and $\theta$ are mutually independent. Figure 6.3 illustrates this setting in terms of a graphical model, where $\theta$ and $\Xi$ appear only in strategy ***S2*** of Section 6.6.
For a more compact notation, a representative variable, $\zeta \in \mathcal{C} = \{k_w, \lambda_w, \lambda_\Delta, \sigma_n^2\}$, is defined along with positive real-valued parameters, $a_\zeta \in \{a, a'', a_\sigma\}$ and $b_\zeta \in \{a, a'', a_\sigma\}$, that belong to the corresponding density functions in (6.19). Related to the set $\mathcal{C}$, a reduced set $\mathcal{C}_{\setminus\zeta} \subset \mathcal{C}$ is defined, which excludes the respective variable $\zeta$.
For Gibbs sampling, the full conditional distributions for each parameter have to be derived. Since $\mathbf{x}, \sigma_n^2$ and $\theta$ are assumed to be independent, one obtains

$$p(\zeta \,|\, \mathbf{y}, \mathbf{x}, \mathcal{C}_{\setminus\zeta}) \;\propto\; p(\mathbf{y} \,|\, \mathbf{x}, \mathcal{C})\, p(\zeta \,|\, \mathbf{x}, \mathcal{C}_{\setminus\zeta}) \tag{6.20}$$

$$\propto\; p(\mathbf{y} \,|\, \mathbf{x}, \mathcal{C})\, p(\zeta \,|\, \mathcal{C}_{\setminus\zeta})\, \tilde{p}(\mathbf{x} \,|\, \mathcal{C})\,. \tag{6.21}$$

The Inv-$\Gamma$ priors of the variables in $\mathcal{C}_{\setminus k_w}$ are conjugate to the Gaussian likelihood function in (6.3). Hence, the posterior distributions of any $\zeta \in \mathcal{C}_{\setminus k_w}$ can be conveniently determined by

$$\zeta \,|\, \mathbf{y}, \mathbf{x}, \mathcal{C}_{\setminus\zeta} \;\sim\; \text{Inv-}\Gamma(\zeta \,|\, a_\zeta + \tfrac{M}{2},\; b_\zeta + \tfrac{1}{2}\,)\; \tilde{p}(\mathbf{x} \,|\, \mathcal{C}), \tag{6.22}$$

The posterior distribution of $k_w$ is given by

$$k_w \,|\, \mathbf{y}, \mathbf{x}, \mathcal{C}_{\setminus k_w} \;\sim\; p(k_w \,|\, \tilde{a}\,', \tilde{b}\,', \tilde{c}\,')\; \tilde{p}(\mathbf{x} \,|\, \mathcal{C}), \tag{6.23}$$

where $\tilde{a}\,' = a' + N, \quad \tilde{b}\,' = b' + \sum_{i=1}^{N} \log(x_i), \quad \text{and} \quad \tilde{c}\,' = (d')^{k_w} + \sum_{i=1}^{N} x_i^{k_w}$ [190]. Samples of the posterior distributions can be obtained, for example, using HMC for each variable separately.

### 6.5.1 The Hamilton Monte Carlo Method

This subsection is a review of the HMC method adapted to the problem of drawing samples from $\tilde{p}(\mathbf{x} \,|\, \mathbf{y}, \mathcal{C})$.

The ideas and concepts of HMC are taken from [189].

HMC is inspired by Hamiltonian dynamics, which is a concept known from classical mechanics. It is used to describe the trajectory of a physical system in a phase space of generalized coordinates and conjugate momenta. This concept is adopted in HMC, where the sparse coefficients, $x_i,\ i = 1, \ldots, N$, represent the coordinates.
In order to describe the dynamics of the sampling process, the potential and kinetic energy of the 'system' have to be defined. To this end, each coordinate is assigned a corresponding conjugate (auxiliary) momentum variable, $\xi_i$, $i = 1, \ldots, N$. The kinetic energy, $K(\boldsymbol{\xi})$, is a function of the variables $\boldsymbol{\xi} = [\xi_1, \ldots, \xi_N]$. It is often defined as the kinetic energy of independent particles with mass $m_i$ moving in free space [189]:

$$K(\boldsymbol{\xi}) = \sum_{i=1}^{N} \frac{\xi_i^2}{2m_i} \,. \tag{6.24}$$

The potential energy is related to the posterior density to be sampled. It is defined by [189]

$$U(\mathbf{x} \,|\, \mathbf{y}, \mathcal{C}) \;=\; -\log \tilde{p}(\mathbf{x} \,|\, \mathbf{y}, \mathcal{C}) - \log(Z_u) \tag{6.25}$$

with normalization constant $Z_u$. To simplify notations, the variables $\mathbf{y}$ and $\mathcal{C}$ are not explicitly written (they are fixed during the sampling process of $\mathbf{x}$). The sampling dynamics are controlled by the 'Hamiltonian function', which is given by [189]

$$\mathcal{H}(\mathbf{x}, \boldsymbol{\xi}) \;=\; U(\mathbf{x}) \;+\; K(\boldsymbol{\xi}) \,. \tag{6.26}$$

The temporal evolution of of a physical system governed by Hamiltonian dynamics can be determined by solving Hamilton's equation of motion [189]:

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} \;=\; \frac{\partial}{\partial \xi_i} \mathcal{H}(\boldsymbol{\xi}, \mathbf{x}) \;=\; \frac{\xi_i}{m_i} \,, \tag{6.27}$$

$$\frac{\mathrm{d}\xi_i}{\mathrm{d}t} \;=\; -\frac{\partial}{\partial x_i} \mathcal{H}(\boldsymbol{\xi}, \mathbf{x}) \;=\; -\frac{\frac{\partial}{\partial x_i} \tilde{p}(\mathbf{x} \,|\, \mathbf{y}, \mathcal{C})}{\tilde{p}(\mathbf{x} \,|\, \mathbf{y}, \mathcal{C})} \,. \tag{6.28}$$

In every HMC iteration, a proposal for a new sample is obtained by the final points of a finite trajectory, denoted by $(x_i^*, \xi_i^*)$. The obtained sample is accepted or rejected based on a Metropolis update with acceptance probability [189]

$$\mathrm{P}(\text{accept}) \;=\; \min\big(\, 1,\ \exp\big(\, -\mathcal{H}(x_i^*, \xi_i^*) \;+\; \mathcal{H}(x_i, \xi_i) \,\big) \,\big) \,. \tag{6.29}$$

In analogy to classical physics, the canonical joint density of the variables $(\mathbf{x}, \boldsymbol{\xi})$ is given by [189]

$$p(\mathbf{x}, \boldsymbol{\xi}) \;=\; \frac{1}{Z_c} \exp\!\left( -\frac{\mathcal{H}(\mathbf{x}, \boldsymbol{\xi})}{T_{\text{sys}}} \right), \tag{6.30}$$

where $T_{\text{sys}}$ is called the 'system temperature' and $Z_c$ is a normalization constant.
In order to obtain a useful expression in HMC, Equation (6.30) can be adapted by conveniently setting $T_{\text{sys}} = 1$ and $Z_c = Z_u$, i.e.

$$p(\mathbf{x}, \boldsymbol{\xi}) \;=\; \tilde{p}(\mathbf{x} \,|\, \mathbf{y}, \mathcal{C}) \prod_{i=1}^{N} \mathcal{N}(\, \xi_i \,|\, 0, m_i \,) \,. \tag{6.31}$$

The Gaussian density of $\xi_i,\ i = 1, \ldots, N$, is due to the choice of $K(\boldsymbol{\xi})$ in (6.24). The factorized form implies independence of $\mathbf{x}$ and $\boldsymbol{\xi}$, such that the desired marginal posterior density of $\mathbf{x}$ can be conveniently approximated using only the samples of $\mathbf{x}$ and discarding the samples of the auxiliary momentum variables, $\boldsymbol{\xi}$. Algorithmic details and extensions of the HMC method are provided in [189].

## 6.6 Parametric Dictionary Learning Strategies

In this section, two parametric dictionary learning strategies are presented. Both are based on the sparse model in Section 6.4. The dependency of the latent variables is shown in Figure 6.3.
The first strategy, referred to as ***S1***, represents one variant of AE-based (iterative) sparse estimation and dictionary learning [54]. It employs a Monte Carlo EM algorithm [163] and is based on the ideas of hybrid Bayesian inference [164, 165], where the dictionary parameter is assumed to be deterministic. In the second strategy, ***S2***, a full Bayesian approach is pursued to jointly estimate the sparse representation, $\mathbf{x}$, the dictionary parameter, $\theta$, and the noise variance, $\sigma_n^2$. Therefore, $\theta$ is incorporated in the Gibbs sampling procedure of Section 6.5. The computational costs of both strategies are governed by drawing samples of the high-dimensional vector $\mathbf{x}$ in each Gibbs sampling iteration. A summary of the corresponding algorithms is provided at the end of this subsection.

### 6.6.1 Hybrid Bayesian Dictionary Learning (*S1*)

In strategy ***S1***, the dictionary parameters are iteratively estimated using a Monte Carlo EM algorithm, where $d = 1, \ldots, d_{\max}$, denotes the iteration index. To begin, an initial

value of the dictionary parameter, $\theta^{(0)}$, is chosen. In each iteration, the joint samples of $\{\mathbf{x}_l, \mathcal{C}_l\}^{(d)}$, $l = 1, \ldots, L_{\mathrm{MC}}$, are drawn using HMC within Gibbs sampling, according to Section 6.5. Based on the previous value of the dictionary parameter, $\theta^{(d-1)}$, the posterior expectation for all $\zeta \in \mathcal{C}$ is calculated by

$$\hat{\zeta}^{(d)} \;=\; \int_{\mathrm{dom}(\zeta)} \zeta \, p(\zeta \,|\, \mathbf{y}, \hat{\theta}^{(d-1)}) \, \mathrm{d}\zeta \quad \approx \quad \frac{1}{L_{\mathrm{MC}}} \sum_{l=1}^{L_{\mathrm{MC}}} \zeta_l^{(d)} \, p(\zeta_l^{(d)} \,|\, \mathbf{y}, \hat{\theta}^{(d-1)}) \,, \tag{6.32}$$

where $\mathrm{dom}(\zeta)$ is the domain of $\zeta$. Next, the posterior mean of the sparse coefficients, $\hat{\mathbf{x}}^{(d)}$, is computed by

$$\hat{\mathbf{x}}^{(d)} \;=\; \int_{\mathbb{R}_+^N} \mathbf{x} \, p(\mathbf{x} \,|\, \mathbf{y}, \hat{\theta}^{(d-1)}) \, \mathrm{d}\mathbf{x} \quad \approx \quad \frac{1}{L_{\mathrm{MC}}} \sum_{l=1}^{L_{\mathrm{MC}}} \mathbf{x}_l^{(d)} \, p(\mathbf{x}_l^{(d)} \,|\, \mathbf{y}, \hat{\theta}^{(d-1)}) \,. \tag{6.33}$$

The positions of the $K$ largest elements in $\hat{\mathbf{x}}^{(d)}$ indicate the current estimates of the translation coefficients, which are denoted by $\hat{\mathcal{S}}^{(d)}$. For the amplitudes of the significant components, the MAP solution is determined, since their posterior distribution has two modes, one around zero and another dominant mode around some larger value. The approximate MAP solution is calculated by

$$\{\hat{\mathbf{x}}, \hat{\mathcal{C}}\}_{\mathrm{MAP}}^{(d)} \;=\; \arg\max_{\mathbf{x}, \mathcal{C}} \; \log p(\mathbf{x}, \mathcal{C} \,|\, \mathbf{y}, \hat{\theta}^{(d-1)}) \tag{6.34}$$

$$\approx \; \arg\max_{\{\mathbf{x}_j, \mathcal{C}_j\} \in \{\mathbf{x}_l, \mathcal{C}_l\}_{l=1,\ldots,L_{\mathrm{MC}}}^{(d)}} \log p(\{\mathbf{x}_j, \mathcal{C}_j\}^{(d)} \,|\, \mathbf{y}, \hat{\theta}^{(d-1)}). \tag{6.35}$$

Finally, the current estimate of the dictionary parameter, $\hat{\theta}^{(d)}$, is computed. In the E-step of the Monte Carlo EM algorithm, the expected value over $\mathbf{x}, \mathcal{C}$ is determined:

$$\mathbb{E}_{\mathbf{x}, \mathcal{C} \,|\, \mathbf{y}, \theta} \; \log p(\mathbf{y}, \mathbf{x}, \mathcal{C} \,|\, \theta) \;=\; \int_{\mathbb{R}_+^N} \int_{\Psi} \log p(\mathbf{y}, \mathbf{x}, \mathcal{C} \,|\, \theta) \, p(\mathbf{x}, \mathcal{C} \,|\, \mathbf{y}, \theta) \, \mathrm{d}\mathcal{C} \, \mathrm{d}\mathbf{x} \tag{6.36}$$

$$\approx \; \frac{1}{L_{\mathrm{MC}}} \sum_{l=1}^{L_{\mathrm{MC}}} \log p(\mathbf{y}, \{\mathbf{x}_l, \mathcal{C}_l\}^{(d-1)} \,|\, \theta) \quad \triangleq \quad Q(\,\theta \,|\, \hat{\theta}^{(d-1)}) \,, \tag{6.37}$$

where $\Psi$ is the product space formed by the domains of all variables in $\mathcal{C}$. Finally, the function $Q(\,\theta \,|\, \hat{\theta}^{(d-1)})$ is maximized with respect to all values of $\theta$ in $\Theta$, yielding the current estimate of the dictionary parameter:

$$\hat{\theta}^{(d)} \;=\; \arg\max_{\theta \,\in\, \Theta} \; Q(\,\theta \,|\, \theta^{(d-1)}\,) \,. \tag{6.38}$$

#### 6.6.1.1 Initialization of $\theta$ via bisectional search

In order to obtain a proper initial value, $\theta^{(0)}$, a bisectional search is carried out. This can help to avoid local optima in the EM algorithm. Since the sparsity level, $K$, of

the desired representation is assumed to be known, a proper initial value is expected to yield a solution with exactly $K$ significant coefficients. At first, an arbitrary value of $\theta^{(0)}$ can be randomly selected. Then, the set of possible values, $\Theta$, is split into two parts. One part contains all values that are larger than the currently selected one. The other part contains the values that are smaller. Next, a solution for $\mathbf{x}$ is determined and the number of significant components, $s$, is compared to $K$. Since $\theta$ describes the level of localization of the signals, the sparse representation exhibits more than $K$ peaks if $\theta$ is chosen too high. Therefore, if $s > K$, the median of the lower division becomes the next trial of $\theta^{(0)}$. If $s < K$, the median of the upper division is selected. This process continues until a proper value of $\theta^{(0)}$ is found, such that $s = K$. A good initialization leads to a faster convergence of the EM algorithm and reduces the chance of getting stuck in local optima.

### 6.6.2 Bayesian Dictionary Learning (*S2*)

In the strategy ***S2***, $\theta$ is treated as a random variable. To this end, each element $\theta_r \in \Theta$ is assigned a corresponding probability mass, $p_r = p(\theta_r)\,, r = 1, \ldots, R_\Theta$, such that $\sum_{r=1}^{R_\Theta} p(\theta_r) = 1$. Hence, $\theta$ follows a categorical (Cat) distribution over the set $\Theta$. The set of corresponding probability masses is denoted by $\boldsymbol{\Xi} = \{p_1, \ldots, p_{R_\Theta}\}$.
In order to account for uncertainty in the prior assumptions for the probability masses, a conjugate Dirichlet (Dir) prior with parameters $\boldsymbol{\nu} = [\nu_1, \ldots, \nu_{R_\Theta}]^\top$ is assigned to $\boldsymbol{\Xi}$. The prior probability density of $\boldsymbol{\Xi}$ is given by

$$p(\boldsymbol{\Xi}) \;=\; \frac{1}{B(\boldsymbol{\nu})} \prod_{r=1}^{R_\Theta} p_r^{\nu_r}\,, \tag{6.39}$$

where the variables $\nu_r$, $r = 1, \ldots, R_\Theta$, are the number of occurrences of the elements in $\Theta$, and $B(\boldsymbol{\nu})$ signifies the Beta-function. After sampling a certain element $\theta_q \in \Theta$, one count is assigned to that element. When another element is sampled, this count is re-assigned to the newly sampled element. A vector $\check{\mathbf{c}} = [c_1, \ldots, c_{R_\Theta}]^\top \in \mathbb{N}^{R_\Theta}$ is defined to indicate the current sample. Its elements are all zero except for the one corresponding to the currently sampled element in $\Theta$. For example, when some $\theta_q,\ q \in \{1, \ldots, R_\Theta\}$, is sampled, then $c_q = 1$ and $c_r = 0\ \forall\ r \in \{1, \ldots, R_\Theta\} \setminus q$.
Without prior knowledge, all elements in $\Theta$ are assumed to be equally likely, i.e. $p_r = 1/R_\Theta\ \forall\ r = 1, \ldots, R_\Theta$. When each element is assigned a single prior count, i.e. $\nu_r = 1\ \forall\ r = 1, \ldots, R_\Theta$, a new sample can significantly change the posterior distribution. However, the posterior remains almost invariant when a larger number of prior counts is assigned, e.g. $\nu_r = \ 100\ \forall\ r = 1, \ldots, R_\Theta$.

The complete model for strategy ***S2*** is obtained by adding the variables $\theta$ and $\boldsymbol{\Xi}$ to the model in (6.19), i.e.

$$\boldsymbol{\Xi} \sim \mathrm{Dir}(\boldsymbol{\Xi} \,|\, \boldsymbol{\nu}) \tag{6.40}$$

$$\theta \,|\, \boldsymbol{\Xi} \sim \mathrm{Cat}(\theta \,|\, R_\Theta, \boldsymbol{\Xi}) \,. \tag{6.41}$$

They are also included in the Gibbs sampling procedure of Section 6.5. The posterior distributions can be determined by considering the dependencies in Figure 6.3, keeping in mind that $\mathbf{x}, \sigma_n^2$ and $\theta$ are assumed to be mutually independent, according to Section 6.5. Hence,

$$\boldsymbol{\Xi} \,|\, \theta \;=\; \widetilde{\boldsymbol{\Xi}} \sim \mathrm{Dir}(\boldsymbol{\Xi} \,|\, \boldsymbol{\nu} + \check{\mathbf{c}}), \tag{6.42}$$

$$\theta \,|\, \mathbf{y}, \widetilde{\boldsymbol{\Xi}} \sim \mathrm{Cat}(\theta \,|\, R_\Theta, \widetilde{\boldsymbol{\Xi}}). \tag{6.43}$$

**Algorithm:** Sparse estimation and dictionary learning strategies ***S1*** & ***S2***

**Input:** $\mathbf{y}, M, \mathbf{\Phi}, N, L, T_d, \delta t, r(t,\theta), K, L_{\text{MC}}, d_{\max}$

**Output:** $\hat{\mathcal{S}}, \hat{\mathbf{x}}, \hat{\theta}, \hat{\sigma}_n, d, ee$

**Parameters:** $a, a', a'', a_\sigma, b, b', b'', b_\sigma, d', \boldsymbol{\nu}, R_\Theta, \{\theta_r\}_{r=1}^{R_\Theta}$,
internal HMC parameters (c.f. [189, 191]).

**0. Initialize:** $\theta$ at random $\rightarrow \hat{\theta}^{(0)}$ via bisectional search,
$\mathbf{A}(\hat{\theta}^{(0)}), \{\hat{\mathbf{x}}^{(0)}, \hat{\mathcal{C}}^{(0)}\}$ as in (6.19), (***S2***): $d_{\max}{=}1$

**1. for** $d = 1$ to $d_{\max}$ **do**

**2.** **for** $l = 1$ to $L_{\text{MC}}$ **do**

**3.** Gibbs sampling: (i) $\mathcal{C}_l^{(d)}$ using (6.22) and (6.23),
(ii) $\mathbf{x}_l^{(d)}$ via HMC.
(***S2***:) (iii) $\theta_l^{(d)}$, $\mathbf{\Xi}_l^{(d)}$using (6.42) and (6.43)

**4.** **end for**

**5.** Estimate: $\hat{\mathcal{S}}^{(d)}$ from $\hat{\mathbf{x}}^{(d)}$ in (6.33),
$\hat{\mathcal{C}}^{(d)}$ from (6.32),
$\hat{\mathbf{x}}_{\text{MAP}}^{(d)}$ from (6.35),

**5.a** (***S1***:) $\hat{\theta}^{(d)} = \arg\max_{\theta\in\Theta}\ Q(\theta \,|\, \hat{\theta}^{(d-1)})$.

**5.b** (***S2***:) $\hat{\theta}^{(d)}$ from (6.32) with $\zeta_l^{(d)} \rightarrow \theta_l^{(d)}$.

**6.** **if** $\hat{\theta}^{(d)} == \hat{\theta}^{(d-1)}$ **or** $d == d_{\max}$

**7.** **return** $\hat{\mathcal{S}}^{(d)}$, $\hat{\mathbf{x}}_{\text{MAP}}^{(d)}$, $\hat{\mathcal{C}}^{(d)}$, $\hat{\theta}^{(d)}$, $ee = \|\mathbf{y} - \mathbf{\Phi}\mathbf{A}(\hat{\theta}^{(d)})\|_2^2$.

**8.** **end if**

**9. end for**

## 6.7 Simulations and Experimental Validation

In this section, the presented sparse estimation and dictionary learning strategies are evaluated in terms of performance and practical applicability for the fiber sensing problem in Chapter 5. To begin, a visualization of the algorithms is provided, showing their qualitative behavior. Next, simulations are used for a quantitative performance analysis, where various scenarios of different SNRs, CS sampling matrices and sample sizes are considered. The results obtained by the presented probabilistic sparse model are compared with those obtained by a deterministic sparse model, using the PDL-OIAI algorithm in Chapter 5. Finally, the strategies ***S1*** and ***S2*** are applied to real fiber sensor data.

### 6.7.1 Simulation Setup

The simulation setup is specified for the fiber sensing problem in Chapter 5.
In this setting, the received localized signals are the temporal reflections from $K = 3$ uniform FBGs. Two reflections are closely spaced and all reflections are modeled with the same amplitude, $A_x$. The task is to determine the delays of the reflections, i.e. the translation coefficients, which are indicated by the index positions of the $K$ most significant components in $\mathbf{x}$. These coefficients are contained in the set $\mathcal{S}$. The dictionary parameter, $\theta$ describes the temporal localization of the received signals in terms of the effective receiver bandwidth. Since this quantity is non-negative, it is assumed that $0 < \theta_{\min}$ and $\theta_{\max} < \infty$. In particular, a number of $R_\Theta = 100$ equally spaced discrete parameter values between 30% and 150% of the true value are considered in the set $\Theta$. For the simulations in this subsection, the dictionary parameter is re-defined relative to its true value, i.e. $\hat{\theta} \rightarrow \hat{\theta}/\theta$. Using this notation, it is easier to see the relative deviation from the true value, which is highly system-specific. A number of $N = 264$ dictionary atoms are generated from $L = 134$ samples of the generating function, $r(t, \theta)$. The delay spacing of the dictionary grid is set to $\delta t = 50$ ns.
The received signal is acquired using CS with two types of sampling matrices, $\mathbf{\Phi}$. Their independent and identically distributed entries are drawn from the following distributions:

(a) Gaussian: $\mathcal{N}(0, 1)$,

(b) Database Friendly (DF) [106]: $\{-1, 0, 1\}$ with probabilities $\{\frac{1}{6}, \frac{2}{3}, \frac{1}{6}\}$.

Approximate inference is accomplished according to Section 6.5. For HMC, the software package *Stan* [192] is used. It provides an efficient implementation of the 'No-U-Turn'

variant of the HMC algorithm [191]. In ***S1***, a proper initialization of $\theta$ is obtained using a bisectional search, according to Section 6.6.1.1. A maximum of $d_{\max} = 35$ EM iterations is considered. In ***S2***, a non-informative prior for $\theta$ is considered, setting $p(\theta_r) = 1/R_\Theta$ and $\nu_r = 1 \; \forall \; r = 1, \ldots, R_\Theta$.

### 6.7.2 Visualization and Working Principle

In this subsection, the working principle and qualitative behavior of the algorithms is visualized using an SNR of 20 dB. CS is performed using a Gaussian sampling matrix and only 50% of the original samples. The results are depicted in Figure 6.4.
The MAP solution for $\mathbf{x}$ is depicted in Figure 6.4 (a). The computation is done according to Section 6.5, based on the true value of $\theta$. It can be observed that augmented sparsity is achieved by using the local covariance model in Section 6.4.1. Without this model, bursts of adjacent coefficients with lower amplitudes are selected due to the high dictionary coherence.
Figure 6.4 (b) shows the posterior density of one coefficient in $\mathbf{x}$, which exhibits a sharp peak around zero for a non-significant coefficient. For a significant coefficient, the distribution is multimodal with a small mode around zero and a dominating mode at some larger value.
In the second row of Figure 6.4, the evolution of the Monte Carlo EM algorithm in strategy ***S1*** is illustrated for different EM iterations. Figure 6.4 (c) shows the MAP solutions for $\mathbf{x}$, i.e. $\hat{\mathbf{x}}_{\mathrm{MAP}}^{(d)}$. It is zoomed on the two peaks on the left side to provide a better visualization. For this simulation, an improper choice for the initial value of $\theta$ leads to more than $K$ significant components in the sparse representation. Nevertheless, the underlying sparse representation (black bullets) is gradually identified in subsequent iterations. It can be seen from Figure 6.4 (d), that also the dictionary parameter, $\theta$, approaches the true value after some iterations.
In Figure 6.4 (e), the function $Q(\theta \,|\, \hat{\theta}^{(d-1)})$ of the first iteration in strategy ***S1*** is plotted for a good and bad initial value of $\theta$. For a bad initialization, the peak of $Q(\theta \,|\, \hat{\theta}^{(d-1)})$ appears far from the true value (black line), while it is close for a properly chosen initial value. This observation indicated that a good initialization can improve the convergence speed of the algorithm and reduce the chance of getting stuck in local optima.
Figure 6.4 (f) depicts the prior and a typical posterior distribution of the parameter $\theta$ in strategy ***S2***. Since the prior is non-informative, all values are equally likely but the posterior distribution is highly peaked around the true value. In this plot, the hyperparameters of the Dirichlet distribution for $\boldsymbol{\Xi}$ are set to $\nu_r = 1 \; \forall \; r = 1, \ldots, R_\Theta$.

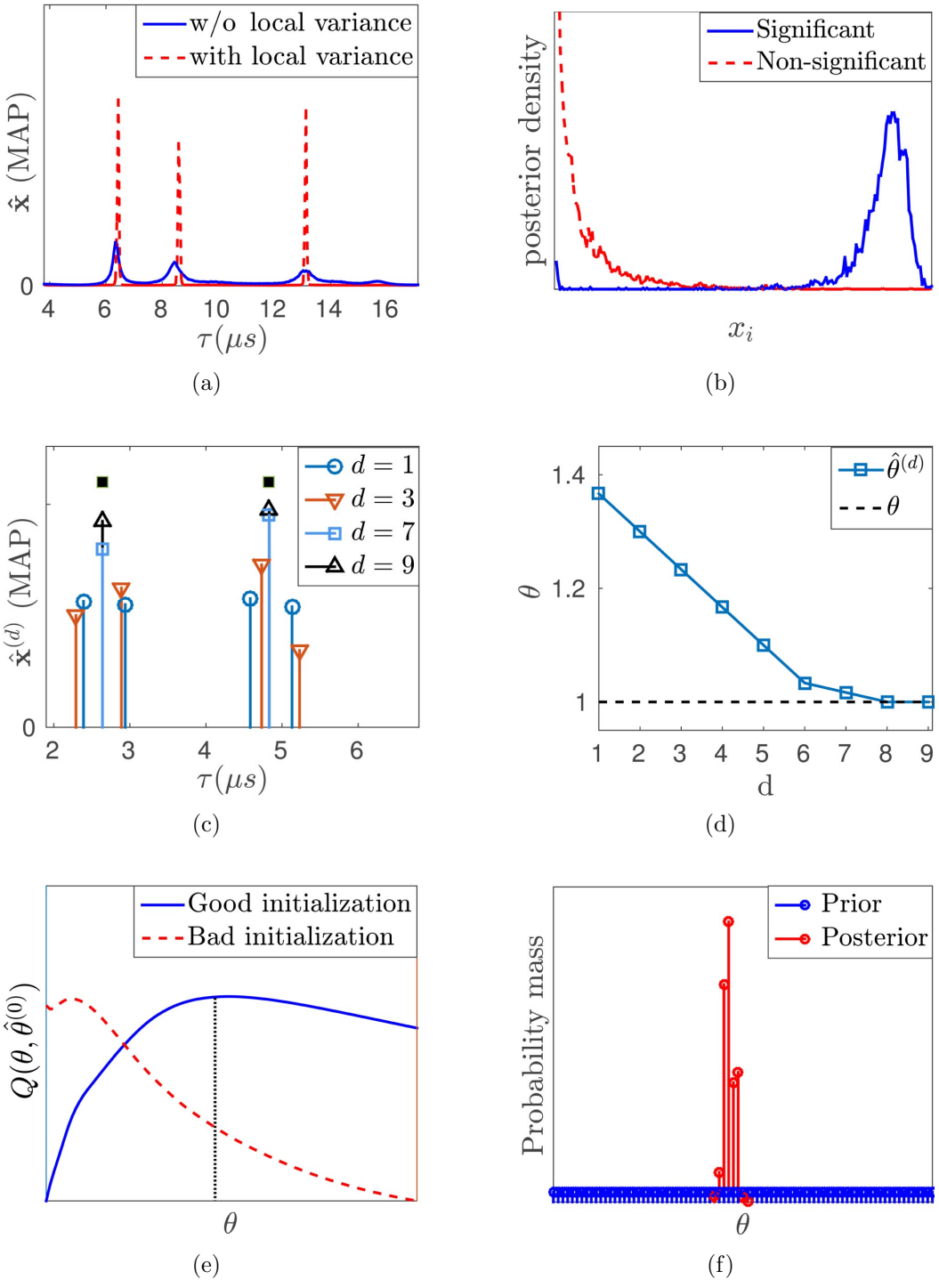



Figure 6.4. Visualization. (a) empirical MAP solution for $\mathbf{x}$ and impact of collective shrinkage, (b) empirical posterior density of a non-/significant entry in $\mathbf{x}$, (c) temporal solution for $\theta$ after the M-step in ***S1***, (d) shape of the $Q$-function in ***S1*** for a good and bad initial value $\theta^{(0)}$, (e) prior and posterior distribution of $\theta$ for $p(\theta_r)=1/R_\Theta$ and $\nu_r=1$, $r=1,\ldots,R_\Theta$ for ***S2***.

### 6.7.3 Performance Evaluation

In this subsection, the performance is evaluated based on simulation. To this end, the *root mean-squared error* (RMSE) is calculated. Given a vector $\mathbf{v}$ and an estimator $\hat{\mathbf{v}}$, the RMSE is defined by

$$\mathrm{RMSE}(\mathbf{v}, \hat{\mathbf{v}}) = \sqrt{\mathbb{E}_{\hat{\mathbf{v}}} \, \|\mathbf{v} - \hat{\mathbf{v}}\|_2^2}\,. \tag{6.44}$$

The expectation is approximated by averaging the results over 100 Monte Carlo runs. The result is denoted by $\overline{\mathrm{RMSE}}$. The strategies ***S1***, ***S2*** are compared to the PDL-OIAI algorithm in Chapter 5, where a deterministic sparse model is assumed. Also, the CRB of Chapter 5 is computed, which is valid for jointly estimating deterministic parameters $(\mathbf{x}_{\mathcal{S}}, \theta)$. It is shown in order to assess a possible performance gain achieved by a probabilistic model. The lower bound of the RMSE induced by the CRB is denoted by 'RCRB'. The results are depicted in Figures 6.5 - 6.7.
The vector $\mathbf{s} \in \mathbb{N}^K$ is defined to contain all elements in $\mathcal{S}$, while $\mathbf{x}_{\mathcal{S}} \in \mathbb{R}_+^K$ is defined to contain the coefficients of $\mathbf{x}$ at the index positions in $\mathcal{S}$. Based on these definitions, $\overline{\mathrm{RMSE}}(\mathbf{x}_{\mathcal{S}}, \hat{\mathbf{x}}_{\mathcal{S}})$ is the error between the estimated amplitudes in $\hat{\mathbf{x}}_{\mathcal{S}}$ at the positions in $\hat{\mathcal{S}}$ and the true amplitudes in $\mathbf{x}_{\mathcal{S}}$ at the true positions in $\mathcal{S}$, which are equally set to the value $A_x$.
The results of strategy ***S1*** for $M/L = 50\%$ of the original samples are shown in Figure 6.5. Figures 6.6 - 6.7, show the results of ***S2***, using 50% and 30% of the original samples, respectively. It should be mentioned that the Monte Carlo EM algorithm in ***S1*** becomes unstable when the ratio M/L is smaller than 50%. Therefore, only this case is shown for ***S1***. Based on this observation, ***S2*** seems more robust against small sample sizes and missing data than ***S1***. For both strategies, and similar as for the PDL-OIAI algorithm in Chapter 5, the choice of the CS sampling matrix, i.e. Gaussian or DF, does not significantly affect the errors. In all scenarios, ***S1*** and ***S2*** are able to estimate $\mathbf{s}$ with significantly higher accuracy than PDL-OIAI. However, PDL-OIAI achieves slightly higher accuracy in estimating the parameter $\theta$ for $M/L = 50\%$, where the error is close to the RCRB. At high SNRs, the performance of ***S2*** is better than that of ***S1***, while ***S1*** is slightly better at lower SNRs. All methods show similar performance for estimating the amplitudes of the sparse coefficients, $\mathbf{x}_{\mathcal{S}}$, where the distance to the RCRB is almost constant at all SNRs. For the noise variance, $\sigma_n^2$, ***S1*** achieves a slightly smaller estimation error as compared to ***S2***. In the case of ***S1*** in Figure 6.5, the RMSE for $\sigma_n^2$ is unstable between SNRs of 15 and 17.5 dB, which might be due to an insufficient number of samples or due to a bad initialization, causing a longer convergence time of the MCMC algorithm to the stationary distribution. In PDL-OIAI, the noise variance cannot be directly estimated.

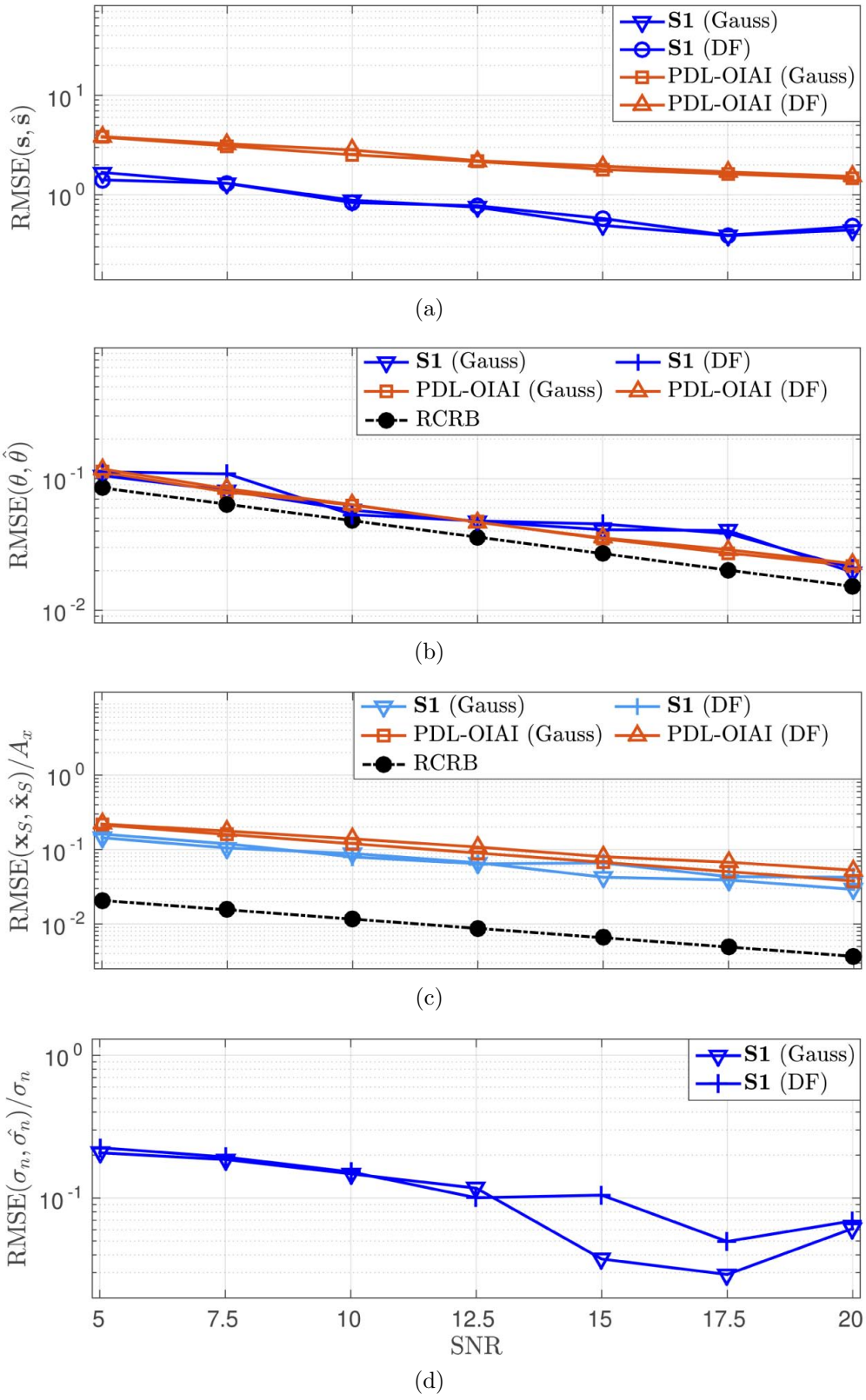



Figure 6.5. Performance of ***S1*** in terms of the RMSE using $M/L = 50\%$ of the original samples in comparison with PDL-OIAI of Chapter 5 and with the lower bound of the RMSE imposed by the CRB (RCRB).

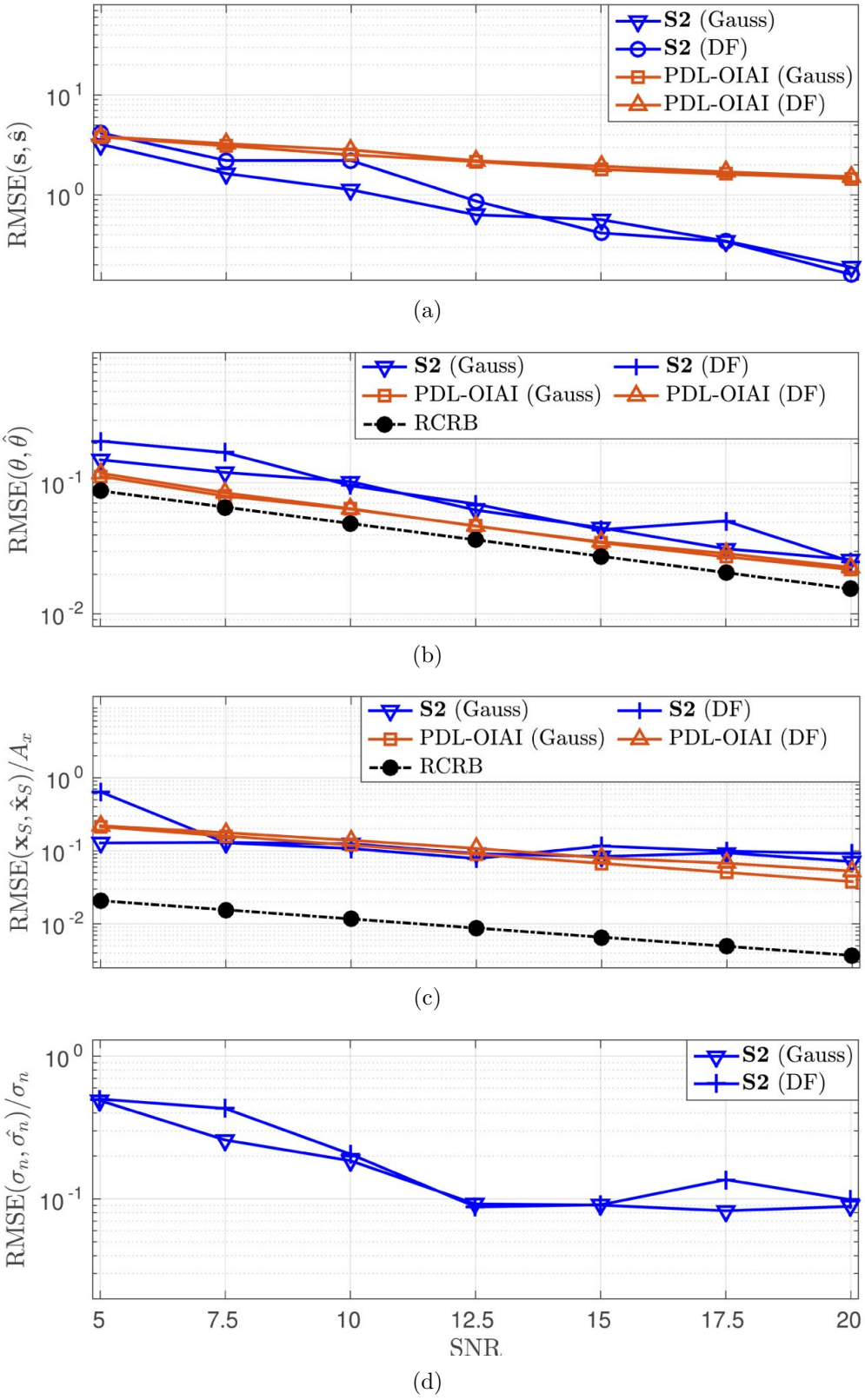



Figure 6.6. Performance of ***S2*** in terms of the RMSE using $M/L = 50\%$ of the original samples in comparison with PDL-OIAI of Chapter 5 and with the lower bound of the RMSE imposed by the CRB (RCRB).

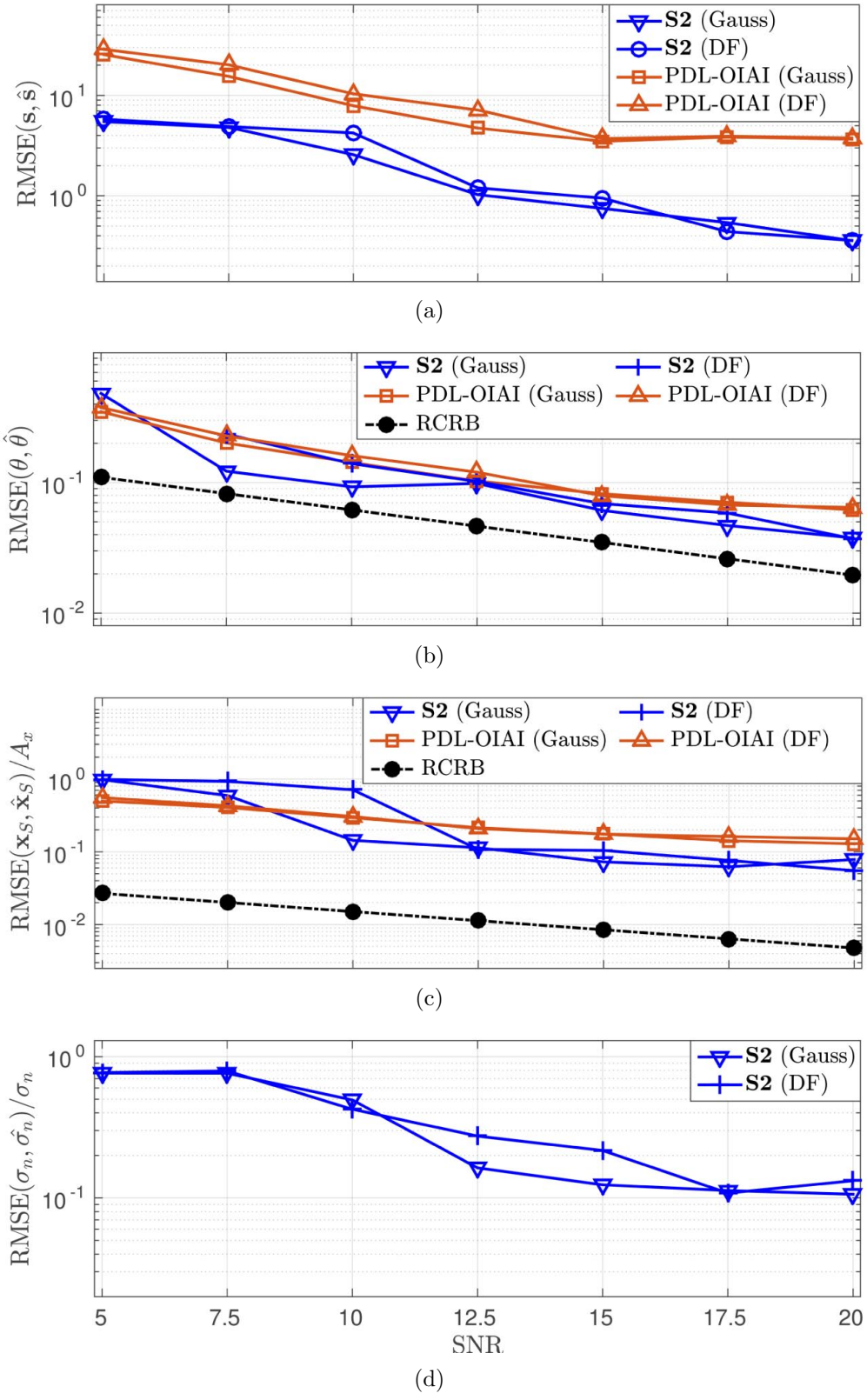



Figure 6.7. Performance of ***S2*** in terms of the RMSE using $M/L = 30\%$ of the original samples in comparison with PDL-OIAI of Chapter 5 and with the lower bound of the RMSE imposed by the CRB (RCRB).

### 6.7.4 Experimental Validation Using Real Data

In this subsection, **S1** and **S2** are applied to real fiber sensing data taken from the fiber sensor in [70,86]. The data was acquired at the Yamashita laboratory of photonic communication devices at The University of Tokyo, Japan.
A number of $L = 134$ original samples of the received sensor signal are used, where only $M/L = 50\%$ samples are acquired using CS. The dictionary contains $N = 2L$ elements and the spacing of the delays described by adjacent dictionary atoms is $\delta t \approx 50$ ns. In the experimental setup, the sensing fiber contains $K = 4$ FBGs. The received reflections are potentially off-grid. By eye inspection of the raw data, their delays are approximately found at [7.79, 9.05, 10.27, 12.30] $\mu$s. The variables $\{\mathcal{S}, \mathbf{x}_\mathcal{S}, \theta\}$ are estimated in 100 Monte Carlo trials.
Figure 6.8 (a) depicts the sensor signal and an estimated reflection from the third FBG. The shaded area signifies the standard deviation of $\theta$. The estimated reflections obtained by ***S2*** are slightly broader, which also results in slightly different estimates of $\mathcal{S}$. Figure 6.8 (b) depicts the estimated sparse representation, i.e. $\hat{\mathbf{x}}_{\hat{\mathcal{S}}}$ evaluated at the estimated positions in $\hat{\mathcal{S}}$. The standard deviation of the amplitudes of the significant components in $\mathbf{x}_\mathcal{S}$ is indicated by vertical error bars, while the standard deviation of $\mathcal{S}$ is illustrated by the shaded areas.
Both strategies, **S1** and **S2**, yield comparable results in this example, although the standard deviation of $\mathcal{S}$ is slightly smaller in the case of **S2**. Similar performance can be observed for the PDL-OIAI algorithm (c.f. Chapter 5).

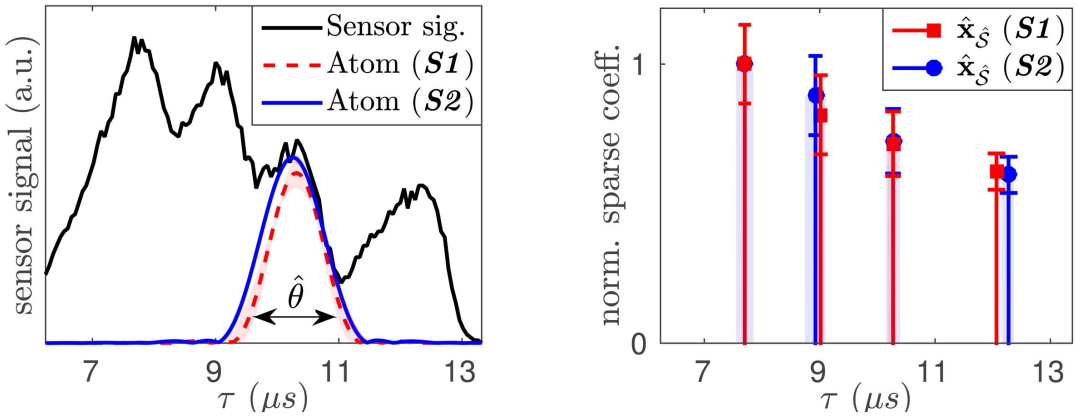



Figure 6.8. Real data example for ***S1*** and ***S2***: (a) sensor signal with estimated reflections. The shaded regions indicate the standard deviation of $\theta$, (b) estimated sparse signal, where the shaded areas show the standard deviation in estimating the reflection delays, indicated by $\mathcal{S}$. The vertical error bars show the standard deviation of $\mathbf{x}_\mathcal{S}$.

## 6.8 Discussion and Findings

Based on simulations and real data, it has been shown that the translation coefficients of localized signals can be accurately estimated using the proposed sparse model in combination with the dictionary learning strategies ***S1*** and ***S2***. The performance is comparable to that of the PDL-OIAI algorithm in Chapter 5, where a deterministic sparse model is considered. Also, the limits imposed by the non-Bayesian CRB for estimating the sparse coefficients, $\mathbf{x}_{\mathcal{S}}$, and the dictionary parameter, $\theta$, are not exceeded. For the task of dictionary learning, all compared methods seem similarly suitable, although ***S2*** and PDL-OIAI are more stable when the sample size is small. However, the primary objective is to estimate the translation coefficients in $\mathbf{s}$ (i.e. the positions of the significant components in $\mathbf{x}$). In fact, the estimation error for $\mathbf{s}$ is significantly lower using the presented probabilistic model, which is a major advantage of this approach. In the problem of fiber sensing, the translation coefficients indicate the reflection delays, which are used to infer the quantity or nature of impairments at the FBGs. The dictionary parameter is estimated to improve the accuracy of the estimates of $\mathbf{s}$, and the amplitudes in $\mathbf{x}_{\mathcal{S}}$ can be used to determine the amount of optical power reflected from the FBGs. In the real data example of this chapter and of Chapter 5, ***S1***, ***S2*** and PDL-OIAI work reliably although some signal features are not captured by the model used to create the dictionary, e.g. the skew shape of the reflections or the signal-dependent noise amplitude.

The results suggest that strong dictionary coherence can be better handled by using a suitable sparse model rather than a dictionary pre-processing routine as in PDL-OIAI. The selective shrinkage properties of the presented sparse model are supported by the local covariance model, which promotes additional collective shrinkage effects. Based on the relation of this model to non-convex optimization with $\ell_p$-norm constraints ($0 < p < 1$), the results further support the findings in [161, 182], which report that, compared to $\ell_1$-minimization, the RIP conditions can be relaxed when non-convex optimization methods are used.

It is further shown that the type of the CS sampling matrix does not significantly impact the results. However, DF matrices [106] are advantageous due to the low storage requirements and the possibility to reduce the average sampling rate.

The computational complexity of ***S1*** and ***S2*** is dominated by drawing samples from the posterior of $\mathbf{x}$ using HMC. Also, their performance depends on the MCMC runtime. For ***S1***, local optima are an additional problem, which can be alleviated using a proper initialization based on a bisectional search, according to Section 6.6.1.1. A good initialization also speeds up the convergence time. Although HMC is highly efficient in sampling high-dimensional parameter spaces in the presence of correlation, MCMC methods are often slower than optimization-based methods [162]. However,

optimization-based methods usually require a regularization parameter to be determined beforehand (e.g. $\ell_1$-minimization) [162]. Greedy methods such as 'Orthogonal Matching Pursuit' [7], in turn, are usually faster than optimization-based methods [25]. Although PDL-OIAI is based on a greedy algorithm, the complexity is dominated by the dictionary pre-processing routine, which is necessary in order to deal with high dictionary coherence. Yet, it can be implemented using parallel processing to speed up computations.
In an extension of this work, multiple CS measurements can be jointly taken into account in order to improve the signal-to-noise ratio and the accuracy. This concept has been used in previous work, e.g. in [5]. Regarding real fiber sensing data, for example, the reflections of the different FBGs are not uniform. Therefore, additional local dictionary parameters can be taken into account to describe particular features of the individual localized signals.

## 6.9 Conclusion

In this chapter, a sparse estimation and dictionary learning framework based on a probabilistic hierarchical sparse model is presented. The results obtained using the probabilistic sparse model are also compared to the results obtained by the PDL-OIAI algorithm in Chapter 5, where a deterministic sparse model is considered. This analysis highlights advantages, disadvantages and limitations of both approaches in order to provide a practical guide to choosing an appropriate model for a specific problem. In addition, the presented framework is applied to the problem of compressed fiber sensing, where real fiber sensing data is used to verify the practical applicability.
In the sparse model, selective shrinkage is achieved using a Weibull prior. This choice is related to non-convex optimization with $\ell_p$-norm constraints. The problem of high dictionary coherence is addressed by leveraging the structure of the dictionary and the signal. To this end, a local covariance model for the sparse coefficients is developed, such that their relation can be described in terms of Markov Random Fields. Based on this model, additional collective shrinkage and augmented sparsity is achieved. This model can be generally used for estimating the translation coefficients of localized signals from highly redundant dictionaries without dictionary pre-processing.
For parametric dictionary learning, a hybrid (***S1***) and a full Bayesian strategy (***S2***) are presented. They are based on either a deterministic or a probabilistic model for the dictionary parameter, respectively. In the deterministic case, a Monte Carlo EM algorithm is used to iteratively estimate this parameter. In the probabilistic case, it is jointly estimated along with the sparse coefficients. Approximate inference is accomplished using a hybrid Markov Chain Monte Carlo method that combines Hamilton

Monte Carlo and Gibbs sampling.
A limitation of the presented framework is the computation time of the employed sampling methods for approximate inference and the difficulty in the diagnosis of Markov Chain convergence and sample independence. The problem of local optima in the iterative estimation strategy, ***S1***, can be alleviated by choosing a proper initial value based on a bisectional search.
In a future analysis, the joint treatment of multiple CS sample vectors can be investigated and additional local dictionary parameters can be taken into account. To reduce the computational complexity, variational Bayes methods can be developed.

# Chapter 7

# Conclusion and Future Directions

In this dissertation, the problem of estimating sparse representations in the presence of uncertainty and errors in the underlying model of the dictionary is addressed. The abstract problems in the theory of sparse representations are analyzed for two representative applications, namely direction finding and compressed fiber-optic sensing. This chapter provides a summary of the developed methodologies and highlights the main conclusions. On this basis, some future research directions are identified and outlined.

## 7.1 Summary and Conclusions

In the first part, a statistical sparse regularization framework for optimization-based sparse estimation is developed. Based on statistical arguments, it offers a systematic strategy for determining a suitable regularization parameter in the presence of model errors and measurement noise. Improved accuracy and stability of the sparse representation as well as robustness against variations in the regularization parameter are achieved by dictionary calibration, which is performed prior to sparse estimation. The presented methods are applied to the problem of direction finding. Simulations are used to evaluate their performance in comparison with existing methods.
In the second part, two methodologies for estimating sparse representations from highly redundant dictionaries are investigated, namely *alternating estimation* (AE) and full Bayesian inference. Uncertainty is considered in terms of unknown dictionary parameters, which are estimated along with the sparse representation. The presented methods differ in the assumption of either a deterministic or a probabilistic sparse model, in the pursued estimation strategies, and in the approaches of dealing with redundancy. In the AE-based methods, the dictionary parameters and the sparse representation are iteratively estimated, where a local cost function is minimized in order to successively improve the estimates. The presented Bayesian framework allows for jointly estimating both quantities. In the case of a deterministic model for the sparse coefficients, the problem of redundancy is addressed by directly modifying the dictionary in order to reduce the similarity between its elements and to avoid ambiguous sparse representations. In the probabilistic case, no prior modification of the dictionary is performed. Instead, a particular sparse model is developed to deal with high redundancy in the dictionary. Both methodologies are applied to a problem in compressed fiber-optic

sensing. Based on simulations, their performance is compared to each other and to theoretical performance bounds. Finally, their practical applicability is validated using real data.

### 7.1.1 Statistical Sparse Regularization Framework

The presented sparse estimation framework yields a suitable regularization parameter for sparse estimation in the presence of model errors. This parameter is estimated as an upper bound of the mean-squared error between the corrupted data and the ideal model. Model errors are considered in terms of a general additive error term that has to be evaluated for the application at hand. Using the direct relation to the considered model error statistics, the regularization parameter can be controlled to account for the specific impairments at hand. In the literature, a good value of the regularization parameter is often determined based on the discrepancy principle, i.e. calculating the sparse solution for a large range of parameters and choosing the one that yields a best-match solution to the observed data. The presented framework yields a suitable and statistically justified regularization parameter prior to sparse estimation, which can reduce the computational complexity. It is applied to the problem of direction finding, where model errors arise as general phase and gain mismatches in the analytic model used to generate the sparsity-promoting dictionary. Simulations show that the tolerance level can be correctly adjusted to allow for model mismatches due to the joint effects of measurement noise and model errors.
A drawback is the the sensitivity of the sparse representation to changes in the regularization parameter and the need for estimating the number of sources, the noise variance and the source power. Also, the choice of the confidence levels for noise and model errors impacts the sparse representation. In order alleviate this problem, dictionary calibration is performed to account for model errors prior to sparse estimation. It only requires a good estimate of the noise subspace. Then, the obtained sparse representation is more accurate and stable, even for large changes of the regularization parameter due to different confidence levels, which is validated using simulations.

### 7.1.2 Sparse Estimation and Dictionary Learning Framework for Fiber-Optic Sensing

A compressed sensing and dictionary learning framework for WDM-based quasi-distributed fiber-optic sensing is presented. Fiber interrogation is performed using a wavelength-tunable laser and the task is to estimated the delays of the reflections.

It is shown that the average sampling rate and the number of samples to be stored and processed can be significantly reduced using compressed sensing-based signal acquisition. Also, it is outlined how a parallel architecture can be used to practically realize the non-uniform sampling scheme described by efficient discrete-valued sparse sampling matrices such as the *Database-Friendly* (DF) matrix.
In order to create a parametric dictionary for estimating the reflection delays, a generic mathematical model for the received sensor signal is compiled and validated using real sensor data. This model can be customized to match different system configurations. For high-accuracy sensing, the dictionary is designed to contain a dense grid of delays, which results in strong similarity, hence, high redundancy. Uncertainty is considered in terms of global and local parameters and reflects imperfect prior knowledge of the signal. In the context of quasi-distributed fiber sensing, local parameters can account for the local characteristics of the fiber Bragg gratings, where the interrogation signal is reflected. Global parameters can account for general characteristics of the system that have common impact on all reflections, e.g. the overall receiver bandwidth.
In order to measure the redundancy level of the employed translation-invariant dictionary the *coherence distance* is introduced. This measure is sensitive to variations in the overall coherence for small sparsity levels. It is a suitable measure for structured dictionaries in which directly adjacent elements exhibit the highest similarity. This is usually the case for translation-invariant dictionaries, where the elements are created from shifted versions of a generating function. Therefore, the coherence distance can be used for general translation-invariant dictionaries. Also, a relation to common coherence measures is established.
For sparse estimation, a deterministic model for the dictionary parameters and the coefficients of the sparse representation is considered. Based on the general formulation of the considered sparse estimation problem, the Cramér-Rao bound for jointly estimating the underlying sparse representation along with the dictionary parameter is derived.
Sparse estimation is performed using an AE-based algorithm, called PDL-OIAI. The problem of redundancy in the dictionary is addressed by adopting a dictionary preprocessing sub-routine that yields a modified sensing dictionary with lower coherence distance. Using this dictionary, a simple greedy algorithm can be used for sparse estimation rather than solving a convex optimization problem. This reduces the computational costs, since sparse estimation has to be performed in each AE iteration. Despite this advantage, it is still costly to calculate the modified sensing dictionary. As a remedy, it can be calculated only once in each AE iteration. The total computational complexity is analyzed for both variants of this method. An additional speed-up can be achieved by parallel computing. Notably, the presented algorithm has broader applicability for general AE-based sparse estimation and dictionary learning problems

with highly redundant dictionaries.
The performance of the presented method is evaluated using simulations, where the estimation error is also compared to the derived Cramér-Rao bound. In comparison to a reference method without dictionary pre-processing (PDL-OMP), it is pointed out that the pre-processing sub-routine in AE-based sparse estimation and dictionary learning is indeed necessary when the dictionary is highly redundant. It is shown that the presented algorithm can accurately estimate the reflection delays and dictionary parameter for moderate coherence levels. The practical applicability is validated using real sensor data.
A general limitation of AE-based estimation is the restriction to local optima in each AE iteration. Also, the computational complexity can be high when many AE iterations have to be carried out.

### 7.1.3 Probabilistic Model for Sparse Estimation of Localized Signals

A Bayesian framework for sparse estimation and dictionary learning with highly redundant translation-invariant dictionaries is developed. The translation-invariant structure of the dictionary is used to estimate localized signals. In order to deal with high redundancy, a probabilistic model for the coefficients of the sparse representation is presented. Besides using a sparsity-promoting prior with favorable selective shrinkage effects, the translation-invariant structure of the dictionary is exploited to achieve augmented sparsity. To this end, a kernel function is introduced, such that the dependency of the sparse coefficients can be described by a Markov Random Field model. Using this model, additional collective shrinkage is achieved, while significant coefficients are allowed to become sufficiently large. Also, a connection between the developed sparse model and non-convex optimization based on the $\ell_p$-norm, $0 < p < 1$, is provided. Previous work has shown that non-convex optimization methods can be superior to convex optimization when the dictionary is highly redundant [161, 182]. The presented sparse model is applied to the fiber sensing problem in Section 7.1.2 but it has broader applicability to the general problem of estimating the translation coefficients of localized signals from highly redundant translation-invariant dictionaries. Notably, redundancy is completely handled by the sparse model and there is no need for any dictionary pre-processing.
For dictionary learning, two strategies are pursued. In the first strategy, a deterministic model for the dictionary parameters is considered. They are estimated using a Monte Carlo Expectation Maximization algorithm as one variant of alternating estimation. A bisectional search is proposed to initialize the algorithm. The second strategy considers

a probabilistic model for the dictionary parameters, yielding a full Bayesian framework. In order to account for uncertainty in prior assumptions, a hierarchical model structure for the sparse coefficients, the dictionary parameters, and the noise variance is considered. A hybrid *Markov Chain Monte Carlo* (MCMC) algorithm is used to accomplish approximate inference in this model. It combines the *Hamilton Monte Carlo* (HMC) method and Gibbs sampling. HMC is highly efficient in high-dimensional spaces and in the presence of correlation. Therefore, it is used to infer the coefficients of the sparse representation. Gibbs sampling is used for all other variables.
The performance is evaluated for the fiber-optic sensing in comparison with the PDL-OIAI method in Section 7.1.2, where a deterministic sparse model is assumed. This comparative analysis can serve as a practical guideline for choosing an appropriate model. Simulations show that the support of the sparse coefficients is more accurately estimated using the Bayesian framework, while the performance for the amplitudes of the coefficients and the dictionary parameters are similar. The Bayesian framework also allows for estimating the noise variance. Its practical applicability is validated using real fiber sensing data. Herein, the Bayesian framework and PDL-OIAI show comparable results. Since MCMC is usually slower than optimization-based or greedy methods, it is also slower than PDL-OIAI. Nevertheless, due to the higher accuracy in estimating the sparse support, the Bayesian framework can be advantageous e.g. for system calibration. Therefore, a combination of both methods is promising.
A drawback of the Bayesian framework is the lack of an analytic solution, such that approximate inference methods have to be applied. Common problems of MCMC methods are the diagnosis of Markov Chain convergence, the sample correlation, and the run-time / computational costs. When the dictionary parameters are estimated using AE-based estimation, the problem of local optima arises. To alleviate this problem, the bisectional search method is used to obtain a good initialization of the algorithm.

## 7.2 Future Research Directions

On the basis of this work, some open problems have been identified, which are stated below.

### 7.2.1 Sparse regularization and dictionary calibration

- Sparse estimation methods can be inaccurate when the magnitudes of the sparse coefficients vary significantly. In the context of sparse regularization, this can lead to non-sparse representations (over-fitting) or suppression of certain features

(under-fitting). Sparse estimation methods that deal with this problem would be of high interest in practice.

- Sparse regularization requires an estimate of the number of sources, the noise power, and also the source power. A detailed analysis on how inaccuracies in these estimates affect the performance can be carried out.
- The presented dictionary calibration method for direction finding requires an estimate of the noise subspace, hence, an estimate of the number of sources. Therefore, robust estimation of the number of sources can be incorporated to improve the calibration procedure for practical systems.
- The presented regularization framework can be extended to deal with off-grid sources. Multi-resolution grid refinement according to [5] can be implemented in order to increase the accuracy and to reduce the negative effects of a limited dictionary grid. Also, it can be analyzed if the framework can be extened to include mutual coupling effects.
- The assumption of uncorrelated Gaussian noise is often too strong in many applications. In particular, R-Capon and W-MUSIC are adaptive algorithms that are able to handle colored noise. Therefore, the impact of the violation of this assumption can be investigated and some simulations with colored noise can be included.

### 7.2.2 Sparse estimation and dictionary learning with highly redundant dictionaries

- The presented sparse estimation and dictionary learning methods use a single set of compressed measurements. Multiple measurements of a stationary scene can be considered to improve the signal-to-noise ratio, the estimation performance, and the stability of the sparse representation. This can be achieved by exploiting goup sparsity among all measurements according to [159].
- In order to monitor a non-stationary scene, efficient algorithms for updating the sparse solution have to be developed or adopted, such that subsequent measurements can be efficiently incorporated.
- In the application of fiber-optic sensing, the signal is acquired using compressive sensing. A direct hardware implementation of this sampling scheme can be considered. The resulting fiber sensing system can be used to provide a more comprehensive experimental validation of the proposed sparse estimation and dictionary learning methods.

- An investigation of the convergence properties of the "Alternating Sparse Estimation" algorithm can be carried out.
- When MCMC methods are used for approximate inference, the computational load can be high. As an alternative, variational Bayes methods can be investigated for the considered problem in order to speed up computations.
- The presented weakly sparse model for the coefficients of the sparse representation favors small entries but does not enforce true sparsity. This model can be compared to a strongly sparse model, e.g. a spike-and-slab model according to [162]. A comprehensive comparison can reveal advantages and disadvantages for the considered application and contribute to the general understanding of both approaches.
- It can be analyzed how the various parameters of the setup, i.e. $L$, $N$, $\delta t$, etc. affect the performance of the different methods.
- A comparison to a conventional filter-bank approach can be performed, where the filters are tuned to each possible value of $\theta$ and the delays are estimated by identifying the peaks in the output of the matched filter. However, it would require perfect knowledge of the shape of the reflections to "match" the filter to the signal.
- Redundancy and incoherence of the dictionary are conflicting requirements. A redundant dictionary offers high-resolution estimates and incoherence is required to ensure the succes of sparse estimation. It can be investigated if and how reducing the dictionary coherence by IAI-mitigation affects the estimation accuracy.

# Appendix

## A.1 Compressed Sampling Using Redundant Dictionaries

When redundant dictionaries are used in *compressed sensing* (CS), the correlations between the row-elements in the combined sensing matrix may lead to problems in sparse estimation [27]. This section reviews some details of this problem and provides an explicit calculation for a Gaussian CS matrix and a general redundant dictionary.

Let $\mathbf{\Phi} = [\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_M]^\top$ be a Gaussian random matrix with independent and identically distributed (i.i.d.) zero-mean entries of variance $\sigma^2$ for which the RIP requirements are fulfilled. Further, let $\mathbf{D} = [\mathbf{d}_1, \ldots, \mathbf{d}_N] \in \mathbb{R}^{L \times N},\ L < N$ be a redundant dictionary. The combined sensing matrix can be written by

$$\mathbf{B} = \mathbf{\Phi D} = \begin{bmatrix} \boldsymbol{\phi}_1^\top \mathbf{d}_1 & \boldsymbol{\phi}_1^\top \mathbf{d}_2 & \ldots & \boldsymbol{\phi}_1^\top \mathbf{d}_N \\ \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{\phi}_M^\top \mathbf{d}_1 & \boldsymbol{\phi}_M^\top \mathbf{d}_2 & \ldots & \boldsymbol{\phi}_M^\top \mathbf{d}_N \end{bmatrix} = \begin{bmatrix} \mathbf{b}_1^\top \\ \vdots \\ \mathbf{b}_M^\top \end{bmatrix}, \tag{1}$$

where $\mathbf{b}_m^\top = \boldsymbol{\phi}_m^\top \mathbf{D},\ m = 1, \ldots, M$. Different rows of $\mathbf{B}$ are independent, since

$$\mathbb{E}_{\mathbf{\Phi}}\ \mathbf{b}_m \mathbf{b}_l^\top \ = \ \mathbb{E}_{\mathbf{\Phi}}\ (\boldsymbol{\phi}_m^\top \mathbf{D})^\top (\boldsymbol{\phi}_l^\top \mathbf{D}) \ = \ \mathbf{D}^\top \left(\mathbb{E}_{\mathbf{\Phi}}\ \boldsymbol{\phi}_m \boldsymbol{\phi}_l^\top\right) \mathbf{D} \ = \ \mathbf{D}^\top \left(\sigma^2 \mathbf{I}\, \delta_{ml}\right) \mathbf{D}\,. \tag{2}$$

Herein, $\mathbb{E}_{\mathbf{\Phi}}\ \mathbf{b}_m \mathbf{b}_l^\top = \mathbf{0}$ for $m \neq l$. However, the correlations between the entries of a certain row do not vanish in case of a redundant dictionary. Their covariance matrix can be calculated by

$$\mathbf{\Sigma} \ = \ \mathbb{E}_{\mathbf{\Phi}}\ \mathbf{b}_m \mathbf{b}_m^\top \ = \ \sigma^2\, \mathbf{D}^\top \mathbf{D} \qquad \forall\ m = 1, \ldots, M. \tag{3}$$

Hence, the off-diagonal elements of $\mathbf{\Sigma}$ are determined by the inter-atom coherence of the dictionary. Hence, the vectors $\mathbf{b}_m^\top$, $m = 1, \ldots, M$, are samples drawn from the distribution $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$ and the correlation between the elements of one row is described by the covariance matrix, $\mathbf{\Sigma}$. Therefore, $\mathbf{B}$ is different from the common type of random matrices used in CS and the RIP results for standard random CS matrices are not applicable. In particular, different $K$-sparse representations, $\mathbf{x}$, may yield the same measurements, $\mathbf{y} = \mathbf{Bx}$. This owes to the fact that the amount of information obtained by a projection with one row of $\mathbf{B}$ is less than that obtained by a random projection with independent and identically distributed Gaussian entries [27]. In other words, even though $\mathbf{\Phi}$ fulfills the RIP requirements, this statement need not be true for $\mathbf{B}$.

## A.2 The Fisher Information: Estimating Parameters in Gaussian Noise

The Fisher information matrix is derived for the case when the parameters, $\boldsymbol{\theta}$, of the dictionary, $\mathbf{D}(\boldsymbol{\theta})$, are perfectly known. For convenience, the dependence on $\boldsymbol{\theta}$ is dropped in $\mathbf{D}(\boldsymbol{\theta})$ and in the likelihood function, $p(\mathbf{y}\,|\,\mathbf{x}, \boldsymbol{\theta})$. Similar derivations for a Gaussian likelihood can also be found elsewhere, for example in [140].

Let $\boldsymbol{\phi}_m$ denote the $m$-the column of a CS matrix, $\boldsymbol{\Phi}^\top$. For independent and identically distributed Gaussian measurements, $y_m$, $m = 1, \ldots, M$, let $[\boldsymbol{\Delta}_{\text{sc}}^{(\mathbf{x})}]_i$, $i = 1, \ldots, N$, denote the $i$-the element of the score function with respect to (w.r.t.) $\mathbf{x}$. It is given by

$$\begin{aligned}
\boldsymbol{\Delta}_{\text{sc}}^{(\mathbf{x})} &= \nabla_{\mathbf{x}} \log p(\mathbf{y}\,|\,\mathbf{x}) \;=\; \frac{1}{p(\mathbf{y}\,|\,\mathbf{x})} \nabla_{\mathbf{x}}\, p(\mathbf{y}\,|\,\mathbf{x}) \\
&= -\frac{1}{2\sigma_n^2} \nabla_{\mathbf{x}} \left( \mathbf{y}^\top \mathbf{y} - \mathbf{x}^\top \mathbf{D}^\top \boldsymbol{\Phi}^\top \mathbf{y} - \mathbf{y}^\top \boldsymbol{\Phi} \mathbf{D} \mathbf{x} + \mathbf{x}^\top \mathbf{D}^\top \boldsymbol{\Phi}^\top \boldsymbol{\Phi} \mathbf{D} \mathbf{x} \right) \\
&= -\frac{1}{2\sigma_n^2} \left( -2\, \mathbf{D}^\top \boldsymbol{\Phi}^\top \mathbf{y} + \mathbf{D}^\top \boldsymbol{\Phi}^\top \boldsymbol{\Phi} \mathbf{D} \mathbf{x} + (\mathbf{D}^\top \boldsymbol{\Phi}^\top \boldsymbol{\Phi} \mathbf{D})^\top \mathbf{x} \right) \\
&= -\frac{1}{2\sigma_n^2} \left( -2\, \mathbf{D}^\top \boldsymbol{\Phi}^\top [\mathbf{y} - \boldsymbol{\Phi} \mathbf{D} \mathbf{x}] \right) \\
&= \frac{1}{\sigma_n^2} \mathbf{D}^\top \boldsymbol{\Phi}^\top (\mathbf{y} - \boldsymbol{\Phi} \mathbf{D} \mathbf{x}) \;. \qquad (4)
\end{aligned}$$

The first moment of the score function is obtained by

$$\mathbb{E}_{\mathbf{y}|\mathbf{x}}\; \boldsymbol{\Delta}_{\text{sc}}^{(\mathbf{x})} = \int \left( \nabla_{\mathbf{x}}\, \log p(\mathbf{y}\,|\,\mathbf{x}) \right)\, p(\mathbf{y}\,|\,\mathbf{x})\, d\mathbf{y} \;= \int \nabla_{\mathbf{x}}\, p(\mathbf{y}\,|\,\mathbf{x})\, d\mathbf{y}. \qquad (5)$$

One usually assumes that integration and differentiation can be exchanged. This assumption holds under some 'weak' regularity conditions, involving the existence and continuity of the partial derivatives w.r.t $\mathbf{x}$, the existence of the integral of $p(\mathbf{x}|\mathbf{y})$ w.r.t. $\mathbf{y}$, and the bounded integrability of the partial derivatives w.r.t. $\mathbf{y}$. Then, the first moment of the score function vanishes, since $\mathbb{E}_{\mathbf{y}|\mathbf{x}}\; \mathbf{y} = \boldsymbol{\Phi}\mathbf{D}\mathbf{x}$:

$$\begin{aligned}
\mathbb{E}_{\mathbf{y}|\mathbf{x}}\; \boldsymbol{\Delta}_{\text{sc}}^{(\mathbf{x})} &= \int \frac{1}{\sigma_n^2} \mathbf{D}^\top \boldsymbol{\Phi}^\top (\mathbf{y} - \boldsymbol{\Phi} \mathbf{D} \mathbf{x})\, p(\mathbf{y}\,|\,\mathbf{x})\, d\mathbf{y} \\
&= \frac{1}{\sigma_n^2} \mathbf{D}^\top \boldsymbol{\Phi}^\top \left( \left( \mathbb{E}_{\mathbf{y}|\mathbf{x}}\; \mathbf{y} \right) - \boldsymbol{\Phi} \mathbf{D} \mathbf{x} \right) \;=\; \mathbf{0}. \qquad (6)
\end{aligned}$$

The second moment can be written in terms of the the score function's Hessian matrix, where $\nabla^2_{\mathbf{x}} := \frac{\partial}{\partial \mathbf{x}} \nabla_{\mathbf{x}}$ yields a matrix. Using the product and the chain rule, one obtains

$$\begin{aligned}
\nabla^2_{\mathbf{x}} \log p(\mathbf{y} \,|\, \mathbf{x}) &= \frac{\partial}{\partial \mathbf{x}} \left( [\nabla_{\mathbf{x}} p(\mathbf{y} \,|\, \mathbf{x})] \, \frac{1}{p(\mathbf{y} \,|\, \mathbf{x})} \right) \\
&= \frac{\nabla^2_{\mathbf{x}} p(\mathbf{y} \,|\, \mathbf{x})}{p(\mathbf{y} \,|\, \mathbf{x})} - \frac{[\nabla_{\mathbf{x}} p(\mathbf{y} \,|\, \mathbf{x})] \cdot [\nabla_{\mathbf{x}} p(\mathbf{y} \,|\, \mathbf{x})]^\top}{p(\mathbf{y} \,|\, \mathbf{x})^2} \\
&= \frac{\nabla^2_{\mathbf{x}} p(\mathbf{y} \,|\, \mathbf{x})}{p(\mathbf{y} \,|\, \mathbf{x})} - \boldsymbol{\Delta}^{(\mathbf{x})}_{\text{sc}} (\boldsymbol{\Delta}^{(\mathbf{x})}_{\text{sc}})^\top \,, \qquad (7)
\end{aligned}$$

where it is assumed that $p(\mathbf{y} \,|\, \mathbf{x})$ and $\nabla_{\mathbf{x}} p(\mathbf{y} \,|\, \mathbf{x})$ are continuous functions in $\mathbf{x}$. Hence, the Fisher information matrix becomes

$$\begin{aligned}
\boldsymbol{\mathcal{I}}(\mathbf{x}) &= \mathbb{E}_{\mathbf{y}|\mathbf{x}} \, \boldsymbol{\Delta}^{(\mathbf{x})}_{\text{sc}} (\boldsymbol{\Delta}^{(\mathbf{x})}_{\text{sc}})^\top \\
&= -\left( \mathbb{E}_{\mathbf{y}|\mathbf{x}} \, \nabla^2_{\mathbf{x}} \log p(\mathbf{y} \,|\, \mathbf{x}) \right) + \int \frac{\nabla^2_{\mathbf{x}} p(\mathbf{y} \,|\, \mathbf{x})}{p(\mathbf{y} \,|\, \mathbf{x})} p(\mathbf{y} \,|\, \mathbf{x}) d\mathbf{y} \\
&= -\left( \mathbb{E}_{\mathbf{y}|\mathbf{x}} \, \nabla^2_{\mathbf{x}} \log p(\mathbf{y} \,|\, \mathbf{x}) \right) + \nabla^2_{\mathbf{x}} \underbrace{\int p(\mathbf{y} \,|\, \mathbf{x}) d\mathbf{y}}_{=1} \qquad (8) \\
&= -\mathbb{E}_{\mathbf{y}|\mathbf{x}} \, \frac{1}{\sigma_n^2} \mathbf{D}^\top \boldsymbol{\Phi}^\top \frac{\partial}{\partial \mathbf{x}} (\mathbf{y} - \boldsymbol{\Phi} \mathbf{D} \mathbf{x}) = \frac{1}{\sigma_n^2} (\boldsymbol{\Phi} \mathbf{D})^\top \boldsymbol{\Phi} \mathbf{D} = \frac{1}{\sigma_n^2} \mathbf{B}^\top \mathbf{B} \,. \qquad (9)
\end{aligned}$$

The final result can also be found, for example, in [139, 140]

# List of Acronyms

| | |
|---|---|
| **AE** | Alternating Estimation |
| **ADC** | Analog-to-Digital Converter |
| **CFS-DL** | Compressed Fiber Sensing and Dictionary Learning |
| **CRB** | Cramér-Rao Bound |
| **CS** | Compressed Sensing |
| **DCF** | Dispersion Compensating Fiber |
| **DF** | Database Friendly |
| **DL** | Dictionary Learning |
| **DOA** | Direction-of-Arrival |
| **DSB** | Delay-and-Sum Beamforming |
| **EM** | Expectation Maximization |
| **FBG** | Fiber Bragg Grating |
| **FIM** | Fisher Information Matrix |
| **HMC** | Hamilton Monte Carlo |
| **IAI** | Inter-Atom-Interference |
| **i.i.d.** | independent and identically distributed |
| **LTI** | Linear Time-Invariant |
| **MAP** | Maximum *a posteriori* |
| **MCMC** | Markov Chain Monte Carlo |
| **MIDR** | Minimum-Interference Distortionless Response |
| **ML** | Maximum Likelihood |
| **MSE** | Mean-Squared Error |
| **OMP** | Orthogonal Matching Pursuit |
| **PD** | Photodetector |

| | |
|---|---|
| **PDL-OIAI** | Parametric DL and OMP-Based Sparse Estimation With IAI Mitigation |
| **PDL-OMP** | Parametric DL and OMP-Based Sparse Estimation |
| **R-CAPON** | Robust Capon Beamforming |
| **RIP** | Restricted Isometry Property |
| **RMSE** | Root Mean-Squared Error |
| **R-SPARSE** | Robust Sparse Regularization Technique |
| **SMF** | Single-Mode Fiber |
| **SNR** | Signal-to-Noise Ratio |
| **SOA** | Semiconductor Optical Amplifier |
| **SPARSE** | Sparse Regularization Technique |
| **WDM** | Wavelength Division Multiplexing |
| **W-MUSIC** | Weighted MUSIC |

# List of Symbols

| | |
|---|---|
| $\mathbf{0}$ | Zero-vector |
| $\mathbf{A}$ | Matrix of array steering vectors |
| $\mathbf{a}$ | Steering vector |
| $B_\epsilon(\mathbf{x}_0)$ | $\epsilon$-ball around some point $\mathbf{x}_0$ |
| $\mathbf{B}$, $\mathbf{b}$ | Combined sensing dictionary, one atom of $\mathbf{B}$ |
| $\mathbf{B}_\mathcal{S}$ | Sub-matrix of $\mathbf{B}$, created from columns with indices in $\mathcal{S}$ |
| $\mathbb{C}$ | Complex numbers |
| $\mathcal{C}$ | Set of variables |
| $\check{\mathbf{c}}$, $c_r$ | Indicator vector, $r$-th entry of $\check{\mathbf{c}}$ |
| $c_p$ | Speed of propagation |
| $\mathbf{D}$, $\mathbf{d}$ | Dictionary, one atom of $\mathbf{D}$ |
| $\hat{\mathbf{D}}$, $\widehat{\Delta\mathbf{d}}$ | Calibrated dictionary, estimated error term |
| $\mathbf{D}'$, $\mathbf{d}'$ | element-wise derivatives of $\mathbf{D}$, $\mathbf{d}$ |
| $\tilde{\mathbf{D}}$, $\tilde{\mathbf{d}}$ | Distorted dictionary, one atom of $\tilde{\mathbf{D}}$ |
| $\Delta\mathbf{D}$, $\Delta\mathbf{d}$ | Error terms corresponding to $\mathbf{D}$, $\mathbf{d}$ |
| $\Delta d$, $\delta d$ | Sensor spacing, spacing error |
| $\mathbf{e}_d$ | $d$-th element in $\mathbf{U}_f$ |
| $f$, $f_0$ | Signal frequencies |
| $f_\mathcal{K}$ | Internal function in $\mathcal{K}$ |
| $\Delta f$, $\widehat{\Delta f}$ | Effective receiver bandwidth, estimate of $\Delta f$ |
| $\mathbf{G}_0$, $\tilde{\mathbf{G}}$ | Ideal sensor gain matrix, distored gain matrix |
| $g_0$ | Ideal sensor gain |
| $\tilde{g}_l$ | Gain error of the $l$-th sensor |
| $\mathbf{I}$ | Identity matrix |
| $I_{\text{ob}}$ | Observation time interval in fiber sensing |

| | |
|---|---|
| $\boldsymbol{\mathcal{I}}$, $\mathcal{I}$ | Fisher information (matrix and scalar version) |
| Inv-$\Gamma$ | Inverse Gamma distribution |
| $J_U$ | Number of internal iteration in PDL-OIAI |
| $K$ | Sparsity level / Number of non-zero elements in $\mathbf{x}$ / Cardinality of $\mathcal{S}$ / Kinetic energy in HMC |
| $\mathcal{K}(\cdot)$ | Kernel function |
| $k_w$ | Shape parameter of the Weibull distribution |
| $\boldsymbol{\mathcal{L}}^2$ | Hilbert space of square-integrable functions |
| $L$ | Number of elements in $\mathbf{r}$ / Number of sensors of an array |
| $L_{\mathrm{MC}}$ | Number of samples in MCMC |
| $M$ | Number of compressed sensing measurements |
| $\mathbf{N}$, $\mathbf{n}$, $n_l$ | Matrix of noise snapshots, noise vector, $l$-th element of $\mathbf{n}$ |
| $\mathbb{N}$, $\mathbb{N}_+$ | Natural numbers, non-negative natural numbers |
| $N$ | Number of elements in $\mathbf{x}$ / Number of atoms in $\mathbf{D}$, $\mathbf{B}$, $\mathbf{W}$ |
| $\hat{n}_{\mathrm{MSE}}$, $\hat{n}_{\mathrm{RMSE}}$, $\hat{q}_{\mathrm{MSE}}$ | Upper bounds for sparse regularization |
| $p_g$, $p_\phi$, $p_e$ | Parameters to control gain and phase errors |
| $\mathbf{Q}$, $\mathbf{q}$ | Error matrix and error vector due to model errors |
| $\tilde{q}_\phi$, $\tilde{q}_\phi$ | Auxiliary random variables |
| $\mathbf{R}$, $\hat{\mathbf{R}}$ | Covariance matrix of sensor snapshots, estimate of $\mathbf{R}$ |
| $\mathbb{R}$, $\mathbb{R}_+$ | Real numbers, non-negative real numbers |
| $R_\Theta$ | Number of parameter values in $\Theta$ |
| $\mathbf{r}$ | Observed or measured signal vector |
| $r(t)$, $r(t, \boldsymbol{\theta})$ | Analog sensor signal, dictionary-generating function with parameters $\boldsymbol{\theta}$ |
| $\mathcal{S}$, $\hat{\mathcal{S}}$ | Support / set of significant elements in $\mathbf{x}$, estimate of $\mathcal{S}$ |
| $\mathbf{s}$ | Vector of elements in $\mathcal{S}$ |
| $\mathbf{T}$ | Linear transformation / Sparse synthesis operator |
| $T$ | Number of snapshots |

| | |
|---|---|
| $T_{\mathrm{sw}}$ | Laser sweep duration |
| $T_d$ | Design sampling rate |
| $T_s$ | Sampling rate |
| $t$ | Time variable |
| $\mathbf{U}_f$ | Matrix of feasible directions |
| $\mathbf{U}_n$, $\hat{\mathbf{U}}_n$ | Noise subspace, estimate of $\mathbf{U}_n$ |
| $\mathbf{U}_{\mathbf{R}}$ | Left-singular vectors with respect to $\mathbf{R}$ |
| $U$ | Potential energy in HMC |
| $u$ | Signal amplitude |
| $\mathbf{V}_{\mathbf{R}}$ | Right-singular vectors with respect to $\mathbf{R}$ |
| $\mathbf{W}$, $\mathbf{w}$ | Modified sensing dictionary, one atom in $\mathbf{W}$ |
| $\mathbf{X}$, $\mathbf{x}$, $x_i$ | Sparse matrix, sparse vector, $i$-th element of $\mathbf{x}$ |
| $\hat{\mathbf{x}}$ | Estimate of $\mathbf{x}$ / Estimator for $\mathbf{x}$ |
| $\hat{\mathbf{x}}_{\mathrm{MAP}}$ | Maximum *a posteriori* estimate of $\mathbf{x}$ |
| $\mathbf{x}_{\mathcal{S}}$ | Sub-vector of $\mathbf{x}$, created from entries with indices in $\mathcal{S}$ |
| $\mathbf{x}^{(\ell_2)}$ | $\ell_2$-norm of one row in $\mathbf{X}$ |
| $\mathbf{Y}$, $\mathbf{y}$, $y_m$ | CS measurement matrix, one vector in $\mathbf{Y}$, $m$-th entry of $\mathbf{y}$ |
| $\mathbf{Z}$ | Covariance matrix of sensor noise |
| $\alpha_{\mathcal{X}}$, $\tilde{\alpha}_{\mathcal{X}}$, $\alpha_\phi$ | Confidence levels |
| $\beta_d$ | Argument of the coherence distance, $d_c(\cdot,\cdot)$ |
| $\beta_{\mathrm{reg}}$, $\hat{\beta}_{\mathrm{reg}}$ | Regularization- / Hyper-parameter, estimate of $\beta_{\mathrm{reg}}$ |
| $\delta_K$ | Restricted isometry constant |
| $\zeta$ | Representative variable in the probabilistic sparse model |
| $\zeta_c$ | Correlation coefficient |
| $\boldsymbol{\eta}$, $\eta$ | Auxiliary momentum variables in HMC, one element of $\boldsymbol{\eta}$ |
| $\Theta$ | Parameter space |

| | |
|---|---|
| $\boldsymbol{\theta}$, $\boldsymbol{\theta}_G$, $\boldsymbol{\theta}_L$, $\theta$ | Dictionary parameter vector, global and local version, one element in $\boldsymbol{\theta}$ |
| $\hat{\boldsymbol{\theta}}$, $\hat{\theta}$ | Estimates of $\hat{\boldsymbol{\theta}}$, $\hat{\theta}$ / Estimators for $\hat{\boldsymbol{\theta}}$, $\hat{\theta}$ |
| $\vartheta$, $\delta\vartheta$ | Angular direction, angular accuracy |
| $\Lambda_B$ | Subset of $\Omega$ |
| $\lambda_B$, $\delta\lambda_B$ | Bragg wavelength, incremental change of $\lambda_B$ |
| $\lambda$, $\lambda_0$ | Signal wavelengths |
| $\lambda_1$, $\lambda_2$ | Hyperparameters for constrained ML estimation |
| $\lambda_\Delta$ | Regularization parameter in $\mathcal{K}(\cdot)$ |
| $\lambda_w$ | Scale parameter of the Weibull distribution |
| $\breve{\mu}_1$, $\breve{\mu}_2$ | Regularization parameters for MAP estimation |
| $\boldsymbol{\nu}$, $\nu_r$ | Dirichlet parameters, $r$-th element of $\boldsymbol{\nu}$ |
| $\boldsymbol{\Xi}$ | Set of probability masses for the elements in $\Theta$ |
| $\boldsymbol{\Sigma}_R$ | Diagonal matrix of singular values of $\mathbf{R}$ |
| $\sigma_g$ | Sensor gain error variance |
| $\sigma_n$, $\hat{\sigma}_n$ | Noise standard deviation, an estimate of $\sigma_n$ |
| $\sigma_s^2$ | Signal power |
| $\sigma_\phi^2$, $\sigma_g^2$ | Phase and gain error variances |
| $\tau$ | Signal delay |
| $\Phi$, $\boldsymbol{\phi}$ | Sampling matrix and measurement vector for CS |
| $\bar{\phi}$, $\Delta\bar{\phi}$ | Measured phase, error in $\bar{\phi}$ |
| $\mathcal{X}^2_{u,(\cdot)}$, $\mathcal{X}^2_{l,(\cdot)}$ | Upper / lower confidence bounds of the $\mathcal{X}^2$-distribution |
| $\mathcal{X}_{u,(\cdot)}$, $\mathcal{X}_{l,(\cdot)}$ | Upper / lower confidence bounds of the $\mathcal{X}$-distribution |
| $\mathcal{N}_{u,\alpha_\phi}$ | Upper confidence bound of the Gaussian distribution |
| $\mathcal{X}_c$ | Locally balanced set |
| $\Omega$ | Index set of dictionary elements |
| $\omega, \omega_0$ | Angular frequencies |

# Functions and Operators

| | |
|---|---|
| $(\cdot)^\top$ | Matrix transposition |
| $(\cdot)^\mathrm{H}$ | Hermitian matrix transposition |
| $(\cdot)^{-1}$ | Matrix inversion |
| $(\cdot)^*$ | Complex conjugation |
| $\|\cdot\|_p$ | $\ell_p$-norm |
| $\|\cdot\|_F$ | Frobenius norm |
| Tr | Trace operator |
| $\frac{\partial f}{\partial \mathbf{x}}$ | Partial derivative of $f$ with respect to $\mathbf{x}$ |
| $\nabla_{(\cdot)}$ | Gradient |
| $\boldsymbol{\Delta}_{\mathrm{sc}}^{(\cdot)}$ | Score function |
| $\prod, \sum, \int$ | Product, sum, integral |
| $\arg\max(\cdot)$ | Arguments of the Maxima |
| $\arg\min(\cdot)$ | Arguments of the Minima |
| $\exp(\cdot)$ | Exponential function |
| $\mathbf{b}(\cdot)$ | Bias function |
| $C(\cdot)$ | Cost function |
| $\mathrm{Cat}(\cdot,\cdot)$ | Categorical probability distribution |
| $\mathcal{CN}(\cdot,\cdot)$ | Complex Gaussian probability density function |
| $\cos(\cdot)$ | Cosine |
| $\det(\cdot)$ | Determinant of a matrix |
| $\mathrm{diag}(\cdot)$ | Diagonal matrix of the arguments |
| $\dim(\cdot)$ | Dimension of a vector space |
| $\mathrm{Dir}(\cdot,\cdot)$ | Dirichlet probability distribution |
| $\mathrm{dom}(\cdot)$ | Domain of a variable |
| $d_c(\cdot,\cdot)$ | Coherence distance of a dictionary |

| | |
|---|---|
| $\mathbb{E}_{\mathbf{y}}$, $\mathbb{E}_{\mathbf{y}\|\mathbf{x}}$ | Expectation with respect to $\mathbf{y}$, conditional expectation given $\mathbf{x}$ |
| $\mathcal{H}(\cdot,\cdot)$ | Hamiltonian function |
| $\mathrm{Im}\{\cdot\}$ | Imaginary part |
| $\log(\cdot)$ | Natural logarithm |
| $\log p(\cdot)$ | Log-probability density function |
| $\mathcal{N}(\cdot,\cdot)$ | Gaussian probability density function |
| $\max(\cdot)$, $\min(\cdot)$ | Maxima, minima |
| $\mathrm{MSE}(\cdot,\cdot)$ | Mean-squared error |
| $\mathcal{O}(\cdot)$ | Big O notation (Bachmann-Landau notation) |
| $\mathrm{P}(\cdot)$ | Probability of the argument |
| $\mathrm{Var}(\cdot)$ | Variance of an estimator |
| $p(\cdot)$, $p(\cdot\,\|\,\cdot)$ | Probability density, conditional probability density |
| $\tilde{p}(\cdot)$, $\tilde{p}(\cdot\,\|\,\cdot)$ | Modified (conditional) probability density |
| $\mathrm{RMSE}(\cdot,\cdot)$ | Root mean-squared error |
| $Q(\cdot,\cdot)$ | Q-function in the Monte Carlo EM algorithm |
| $\mathcal{R}(\cdot)$ | Range space of a matrix |
| $\mathrm{rank}(\cdot)$ | Rank of a matrix |
| $\mathrm{Re}\{\cdot\}$ | Real part |
| $\sin(\cdot)$ | Sine |
| $\mathrm{span}(\cdot)$ | Span of a set of vectors |
| $\mathrm{spark}(\cdot)$ | Spark of a matrix |
| $\mathrm{supp}(\cdot)$ | Support set of a vector |
| $\mathcal{W}(\cdot,\cdot)$ | Weibull probability density function |
| $\delta_{(\cdot,\cdot)}$ | Kronecker's delta function |
| $\mu_{\mathrm{MC}}(\cdot)$, $\mu_B(\cdot)$ | Mutual coherence, $\ell_1$-coherence / Babel function |
| $\mathcal{X}$, $\mathcal{X}^2$ | $\mathcal{X}$-distribution, $\mathcal{X}^2$-distribution |

# Bibliography

[1] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1798 – 1828, Aug 2003.

[2] V. M. Patel and R. Chellappa, *Sparse Representations and Compressive Sensing for Imaging and Vision*, Springer, 2013.

[3] R. Rubinstein, A. M. Bruckstein, and M. Elad, "Dictionaries for sparse representation modeling," *Proc. IEEE*, vol. 98, no. 6, pp. 1045–1057, Jun 2010.

[4] C. Christopoulos, A. Skodras, and T. Ebrahimi, "The JPEG2000 still image coding system: an overview," *IEEE Trans. Consum. Electron.*, vol. 46, no. 4, pp. 1103–1127, Nov 2000.

[5] D. Malioutov, M. Cetin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3010–3022, Aug 2005.

[6] A. Gilbert, S. Muthukrishnan, and M. J. Strauss, "Approximation of functions over redundant dictionaries using coherence," in *Proc. 14th Annu. ACM-SIAM Symp. Discrete algorithms (SODA)*, Portland, Oregon, USA, Jan 2003, pp. 243–252.

[7] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition," in *Proc. 27th Asilomar Conf. Signals Syst. Comput. (ASILOMAR)*, Pacific Grove, California, USA, Nov 1993, vol. 1, pp. 40–44.

[8] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Rev.*, vol. 43, no. 1, pp. 129–159, Jan 2001.

[9] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via $\ell_1$-minimization," *Proc. Nat. Acad. Sci. (PNAS)*, vol. 100, no. 5, pp. 2197–2202, Mar 2003.

[10] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harmon. Anal.*, vol. 26, no. 3, pp. 301–321, Apr 2009.

[11] W. R. Babbitt, Z. W. Barber, and C. Renner, "Compressive laser ranging," *Opt. Lett.*, vol. 36, no. 24, pp. 4794–4796, Dec 2011.

[12] J. A. Tropp, "Greed is good: algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct 2004.

[13] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec 2005.

[14] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online dictionary learning for sparse coding," in *Proc. 26th Annu. Int. Conf. Mach. Learn. (ICML)*, New York, USA, Jun 2009, pp. 689–696.

[15] Z. Jiang, Z. Lin, and L. S. Davis, "Label consistent K-SVD: Learning a discriminative dictionary for recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 11, pp. 2651–2664, Nov 2013.

[16] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly, "Compressed sensing MRI," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 72–82, Mar 2008.

[17] H. Yu and G. Wang, "Compressed sensing based interior tomography," *Phys. Med. Biol.*, vol. 54, no. 9, pp. 2791–2805, May 2009.

[18] V. M. Patel, G. R. Easley, D. M. Healy Jr., and R. Chellappa, "Compressed synthetic aperture radar," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 244–254, Apr 2010.

[19] M. G. Amin, *Through-The-Wall Radar Imaging*, CRC Press, 2010.

[20] R. G. Baraniuk, "Compressive sensing [lecture notes]," *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 118–121, Jul 2007.

[21] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr 2006.

[22] Y. C. Eldar and G. Kutyniok, *Compressed Sensing: Theory and Applications*, Cambridge University Press, 2012.

[23] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*, Springer, 2010.

[24] M. Elad and A. M. Bruckstein, "A generalized uncertainty principle and sparse representation in pairs of bases," *IEEE Trans. Inf. Theory*, vol. 48, no. 9, pp. 2558–2567, Sep 2002.

[25] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*, Birkhäuser, 2013.

[26] E. J. Candes and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar 2008.

[27] E. J. Candes, Y. C. Eldar, D. Needell, and P. Randall, "Compressed sensing with coherent and redundant dictionaries," *Appl. Comput. Harmon. Anal.*, vol. 31, pp. 59–73, Jul 2011.

[28] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 2845–2862, Nov 2001.

[29] K. Schnass and P. Vandergheynst, "Dictionary preconditioning for greedy algorithms," *IEEE Trans. Signal Process.*, vol. 56, no. 5, pp. 1994–2002, May 2008.

[30] C. Weiss and A. M. Zoubir, "A compressed sampling and dictionary learning framework for wavelength-division-multiplexing-based distributed fiber sensing," in *J. Opt. Soc. Am. A (JOSA A)*. 2017, Optical Society of America (OSA), accepted in March 2017 (assigned issue: vol. 34, no. 5). Online available: arXiv:1609.08043v2 [stat.ME]
[©2017 Optical Society of America.]. One print or electronic copy may be made for personal use only. Systematic reproduction and distribution, duplication of any material in this paper for a fee or for commercial purposes, or modifications of the content of this paper are prohibited.

[31] H. Rauhut, K. Schnass, and P. Vandergheynst, "Compressed sensing and redundant dictionaries," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 2210–2219, May 2008.

[32] R. Yang, Q. Wan, Y. Liu, and W. Yang, "Adaptive inter-atom interference mitigation approach to sparse multi-path channel estimation," in *Proc. IEEE 71st Veh. Technol. Conf. (VTC)*, Taipei, Taiwan, May 2010, pp. 1–5.

[33] R. Gribonval and M. Nielsen, "Sparse representations in unions of bases," *IEEE Trans.Inf. Theory*, vol. 49, no. 12, pp. 3320–3325, Dec 2003.

[34] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, Dec 1993.

[35] M. Ataee, H. Zayyani, M. Babaie-Zadeh, and C. Jutten, "Parametric dictionary learning using steepest descent," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Dallas, Texas, Mar 2010, pp. 1978–1981.

[36] P. Jost, P. Vandergheynst, S. Lesage, and R. Gribonval, "Learning redundant dictionaries with translation invariance property: the MoTIF algorithm," in *Proc. Signal Process. Adapt. Sparse Struct. Represent. (SPARS)*, Rennes, France, Jan 2005, pp. 1–3.

[37] P. Jost, P. Vandergheynst, S. Lesage, and R. Gribonval, "MoTIF: An efficient algorithm for learning translation invariant dictionaries," in *Proc. IEEE Int. Conf. Acoust. Speech, Signal Process. (ICASSP)*, Toulouse, France, May 2006, vol. 5, pp. V–V.

[38] T. Blumensath and M. Davies, "Sparse and shift-invariant representations of music," *IEEE Trans. on Audio, Speech, Language Process.*, vol. 14, no. 1, pp. 50–57, Jan 2006.

[39] K. Fyhn, M. F. Duarte, and S. H. Jensen, "Compressive parameter estimation for sparse translation-invariant signals using polar interpolation," *IEEE Trans. Signal Process.*, vol. 63, no. 4, pp. 870–881, Feb 2015.

[40] J. Mairal, M. Elad, and G. Sapiro, "Sparse representation for color image restoration," *IEEE Trans. Image Proc.*, vol. 17, no. 1, pp. 53–69, Jan 2008.

[41] M. Yaghoobi, L. Daudet, and M. E. Davies, "Parametric dictionary design for sparse coding," in *Proc. Signal Process. Adapt. Sparse Struct. Represent. (SPARS)*, R. Gribonval, Ed., Saint Malo, France, Apr 2009.

[42] K. Engan, S. O. Aase, and J. Hakon Husoy, "Method of optimal directions for frame design," in *Proc. IEEE Int. Conf. Acoust. Speech, Signal Process. (ICASSP)*, Phoenix, Arizona, USA, Mar 1999, vol. 5, pp. 2443–2446.

[43] K. Engan, B. D. Rao, and K. Kreutz-Delgado, "Frame design using FOCUSS with method of optimal directions (MOD)," in *Proc. Norwegian Signal Process. Symp. (NORSIG)*, Asker, Norway, 1999, pp. 65–69.

[44] R. Vidal, Y. Ma, and S. Sastry, "Generalized principal component analysis (GPCA)," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 12, pp. 1945–1959, Dec 2005.

[45] K. Engan, K. Skretting, and J. Hakon Husoy, "Family of iterative LS-based dictionary learning algorithms, ILS-DLA, for sparse signal representation," *Digit. Signal Process.*, vol. 17, no. 1, pp. 32–49, Jan 2007.

[46] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov 2006.

[47] K. Skretting and K. Engan, "Recursive least squares dictionary learning algorithm," *IEEE Trans. Signal Process.*, vol. 58, no. 4, pp. 2121–2130, Apr 2010.

[48] Z. Chen and Y. Wu, "Robust dictionary learning by error source decomposition," in *Proc. IEEE Int. Conf. Comput. Vision (ICCV)*, Sydney, Australia, Dec 2013, pp. 2216–2223.

[49] R. Gribonval, R. Jenatton, and F. Bach, "Sparse and spurious: Dictionary learning with noise and outliers," *IEEE Trans. Inf. Theory*, vol. 61, no. 11, pp. 6298–6319, Nov 2015.

[50] M. S. Lewicki and T. J. Sejnowski, "Learning overcomplete representations," *Neural Comput.*, vol. 12, no. 2, pp. 337–365, Feb 2000.

[51] K. Kreutz-Delgado, J. F. Murray, B. D. Rao, K. Engan, T. W. Lee, and T. J. Sejnowski., "Dictionary learning algorithms for sparse representation," *Neural Comput.*, vol. 15, no. 2, pp. 349–396, Feb 2003.

[52] S. Arora, R. Ge, and A. Moitra, "New algorithms for learning incoherent and overcomplete dictionaries," *J. Machine Learn. R.*, vol. 35, pp. 1–28, May 2014.

[53] A. Agarwal, A. Anandkumar, P. Jain, P. Netrapalli, and R. Tandon, "Learning sparsely used overcomplete dictionaries," *J. Mach. Learn. R.*, vol. 35, pp. 1–15, May 2014.

[54] A. Beck and L. Tetruashvili, "On the convergence of block coordinate descent type methods," *SIAM J. Optim.*, vol. 23, no. 4, pp. 2037–2060, Jan 2013.

[55] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. R. Stat. Soc. Series B*, vol. 39, no. 1, pp. 1–38, 1977.

[56] T. L. Hansen, M. A. Badiu, B. H. Fleury, and B. D. Rao, "A sparse Bayesian learning algorithm with dictionary parameter estimation," in *Proc. IEEE 8th Sensor Array and Multichannel Signal Process. Workshop (SAM)*, A Coruna, Spain, Jun 2014, pp. 385–388.

[57] M. Elad, P. Milanfar, and R. Rubinstein, "Analysis versus synthesis in signal priors," *Inverse Probl.*, vol. 23, no. 3, pp. 947–968, Apr 2007.

[58] O. Christensen, *An Introduction to Frames and Riesz Bases*, Springer, 2003.

[59] E. J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Jan 2006.

[60] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. R. Stat. Soc. Series B*, vol. 58, pp. 267–288, 1994.

[61] T. Park and G. Casella, "The bayesian Lasso," *J. Amer. Statist. Assoc.*, vol. 103, no. 482, pp. 681–686, Jun 2008.

[62] E. J. Candes, "The restricted isometry property and its implications for compressed sensing," *C.R. Math.*, vol. 346, pp. 589–592, May 2008.

[63] M. A. Davenport and M. B. Wakin, "Analysis of orthogonal matching pursuit using the restricted isometry property," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4395–4401, Sep 2010.

[64] H. L. Van Trees, *Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory*, John Wiley & Sons, 2002.

[65] H. Krim and M. Viberg, "Two decades of array signal processing research: the parametric approach," *IEEE Signal Process. Mag.*, vol. 13, no. 4, pp. 67–94, Jul 1996.

[66] C. Weiss and A. M. Zoubir, "A sparse regularization technique for source localization with non-uniform sensor gain," in *Proc. IEEE 8th Sensor Array Multichannel Signal Process. Workshop (SAM)*, A Coruna, Spain, Jun 2014, Copyright ©2014 by IEEE. Online: http://ieeexplore.ieee.org/document/6882346/. DOI: 10.1109/SAM.2014.6882346.

[67] C. Weiss and A. M. Zoubir, "DOA estimation in the presence of array imperfections: A sparse regularization parameter selection problem," in *Proc. IEEE Workshop Stat. Signal Process. (SSP)*, Gold Coast, Australia, Jun 2014, Copyright ©2014 by IEEE. Online: http://ieeexplore.ieee.org/document/6884647/. DOI: 10.1109/SSP.2014.6884647.

[68] C. Weiss and A. M. Zoubir, "Robust high-resolution DOA estimation with array pre-calibration," in *Proc. 22nd Eur. Signal Process. Conf. (EUSIPCO)*, Lisbon, Portugal, Sep 2014, Copyright ©2014 by EURASIP. Online: http://ieeexplore.ieee.org/document/6952349/. ISSN: 2076-1465.

[69] K. Hotate and Z. He, "Synthesis of optical-coherence function and its applications in distributed and multiplexed optical sensing," *J. Lightwave Technol.*, vol. 24, no. 7, pp. 2541, Jul 2006.

[70] Y. Nakazaki and S. Yamashita, "Fast and wide tuning range wavelength-swept fiber laser based on dispersion tuning and its application to dynamic FBG sensing," *Opt. Express*, vol. 17, no. 10, pp. 8310–8318, May 2009.

[71] K. O. Hill and G. Meltz, "Fiber bragg grating technology fundamentals and overview," *J. Lightwave Technol.*, vol. 15, no. 8, pp. 1263 –1276, Aug 1997.

[72] T. Amano, H. Hiro-Oka, D. Choi, H. Furukawa, F. Kano, M. Takeda, M. Nakanishi, K. Shimizu, and K. Ohbayashi, "Optical frequency-domain reflectometry with a rapid wavelength-scanning superstructure-grating distributed Bragg reflector laser," *Appl. Opt.*, vol. 44, no. 5, pp. 808–816, Feb 2005.

[73] K. Hung, C. C. Lee, W. M. Chan, S.-O. Choy, and P. Kwok, "Development of novel wearable sensors for mobile health," in *Proc. IEEE-EMBS Int. Conf. on Biomedical and Health Informatics (BHI)*, Hong Kong, China, Jan 2012, pp. 745 –747.

[74] M. Iodice, V. Striano, G. Cappuccino, and G. Cocorullo, "Fiber Bragg grating sensors-based system for strain measurements," in *Proc. 2005 IEEE/LEOS Workshop Fibres Optical Passive Components (WFOPC)*, Palermo, Italy, Jun 2005, pp. 307–312.

[75] E. Udd, "Fiber optic smart structures," *Proc. IEEE*, vol. 84, no. 1, pp. 60–67, Jan 1996.

[76] A. D. Kersey, M. A. Davis, H. J. Patrick, M. LeBlanc, K. P. Koo, C. G. Askins, M. A. Putnam, and E. J. Friebele, "Fiber grating sensors," *J. Lightwave Technol.*, vol. 15, no. 8, pp. 1442–1463, Aug 1997.

[77] A. D. Kersey, "Optical fiber sensors for permanent downwell monitoring applications in the oil and gas industry," *IEICE Trans. Electron.*, vol. E83-C, pp. 400–404, Mar 2000.

[78] R. M. Measures, *Structural Monitoring with Fiber Optic Technology*, Academic Press Inc., 2001.

[79] A. Mendez and T. Morse, *Specialty Optical Fibers Handbook*, Elsevier, 2007.

[80] B. Culshaw and A. Kersey, "Fiber-optic sensing: A historical perspective," *J. Lightwave Technol.*, vol. 26, no. 9, pp. 1064–1078, May 2008.

[81] Y. R. Garcia, J. M. Corres, and J. Goicoechea, "Vibration detection using optical fiber sensors," *J. Sensors*, p. 12 pp, 2010.

[82] B. Culshaw, "Optical fiber sensor technologies: Opportunities and - perhaps - pitfalls," *J. Lightwave Technol.*, vol. 22, no. 1, pp. 39, Jan 2004.

[83] A. Rosenthal and M. Horowitz, "Inverse scattering algorithm for reconstructing lossy fiber Bragg gratings," *J. Opt. Soc. Am. A*, vol. 21, no. 4, pp. 552–560, Apr 2004.

[84] K. Hotate and K. Kajiwara, "Proposal and experimental verification of Bragg wavelength distribution measurement within a long-length FBG by synthesis of optical coherence function," *Opt. Express*, vol. 16, no. 11, pp. 7881–7887, May 2008.

[85] H. Murayama, D. Wada, and H. Igawa, "Structural health monitoring by using fiber-optic distributed strain sensors with high spatial resolution," *J. Photonic Sensors*, vol. 3, no. 4, pp. 355–376, Dec 2013.

[86] S. Yamashita, Y. Nakazaki, R. Konishi, and O. Kusakari, "Wide and fast wavelength-swept fiber laser based on dispersion tuning for dynamic sensing," *J. Sensors*, vol. 2009, pp. 12, 2009.

[87] C. Weiss and A. M. Zoubir, "Fiber sensing using UFWT-lasers and sparse acquisition," in *Proc. 21st Eur. Signal Process. Conf. (EUSIPCO)*, Marrakech, Morocco, Sep 2013, pp. 1–5, Copyright ©2013 by EURASIP. Online: http://ieeexplore.ieee.org/document/6811589/. ISSN: 2076-1465.

[88] C. Weiss and A.M. Zoubir, "Fiber sensing using wavelength-swept lasers: A compressed sampling approach," in *Proc. 3rd Int. Workshop Compressed Sens. Theory Appl. Radar Sonar Remote Sens. (CoSeRa)*, Pisa, Italy, Jun 2015, pp. 21–25, Copyright ©2015 by IEEE. Online: http://ieeexplore.ieee.org/document/7330256/. DOI: 10.1109/CoSeRa.2015.7330256.

[89] T. Erdogan, "Fiber grating spectra," *J. Lightwave Technol.*, vol. 15, no. 8, pp. 1277–1294, Aug 1997.

[90] W. W. Morey, G. Meltz, and W. H. Glenn, "Fiber optic Bragg grating sensors," *Proc. SPIE 1169, Fiber Optic Laser Sensors VII*, vol. 1169, pp. 98, Feb 1990.

[91] B. Culshaw, G. Thursby, D. Betz, and B. Sorazu, "The detection of ultrasound using fiber-optic sensors," *IEEE Sensors J.*, vol. 8, no. 7, pp. 1360–1367, Jul 2008.

[92] A. Othonos and K. Kalli, *Fiber Bragg Gratings: Fundamentals and Applications in Telecommunications and Sensing*, Artech House, 1999.

[93] R. Kashyap, *Fiber Bragg Gratings*, AP, 2nd edition, 2009.

[94] S. Yin, P. B. Ruffin, and F. T. S. Yu, *Fiber Optic Sensors*, CRC Press, 2nd edition, 2008.

[95] M. A. Herman and T. Strohmer, "High-resolution radar via compressed sensing," *IEEE Trans. Signal Process.*, vol. 57, no. 6, pp. 2275–2284, Jun 2009.

[96] N. A. Goodman and L. C. Potter, "Pitfalls and possibilities of radar compressive sensing," *Appl. Opt.*, vol. 54, no. 8, pp. C1–C13, Mar 2015.

[97] M. E. Gehm and D. J. Brady, "Compressive sensing in the EO/IR," *Appl. Opt.*, vol. 54, no. 8, pp. C14–C22, Mar 2015.

[98] Y.-P. Lin and P. P. Vaidyanathan, "Periodically nonuniform sampling of bandpass signals," *IEEE Trans. Circuits Syst. II: Analog Digit. Signal Process.*, vol. 45, no. 3, pp. 340–351, Mar 1998.

[99] M. Mishali and Y.C. Eldar, "Blind multiband signal reconstruction: Compressed sensing for analog signals," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 993–1009, Mar 2009.

[100] J. Sheng, C. Yang, and M. C. Herbordt, "Hardware-efficient compressed sensing encoder designs for WBSNs," in *Proc. IEEE High Perform. Extreme Comput. Conf. (HPEC)*, Waltham, Massachusetts, USA, Sep 2015, pp. 1–7.

[101] Z. Yu, S. Hoyos, and B. M. Sadler, "Mixed-signal parallel compressed sensing and reception for cognitive radio," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Las Vegas, Nevada, USA, Mar 2008, pp. 3861–3864.

[102] T. Strohmer and J. Tanner, "Fast reconstruction algorithms for periodic nonuniform sampling with applications to time-interleaved ADCs," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Honolulu, Hawai'i, USA, Apr 2007, vol. 3, pp. III–881–III–884.

[103] M. Mishali, Y. C. Eldar, and A. J. Elron, "Xampling: Signal acquisition and processing in union of subspaces," *IEEE Trans. Signal Process.*, vol. 59, no. 10, pp. 4719–4734, Oct 2011.

[104] X. Xiong, C. Han, Q. Ding, L. Liu, and D. Li, "A parallel delta-sigma ADC based on compressive sensing," in *Proc. IEEE Int. Conf. Electron. Devices Solid-State Circuits (EDSSC)*, Hong Kong, China, Jun 2013, pp. 1–2.

[105] M. Elad, "Optimized projections for compressed sensing," *IEEE Trans. Signal Process.*, vol. 55, no. 12, pp. 5695–5702, Dec 2007.

[106] D. Achlioptas, "Database-friendly random projections: Johnson-Lindenstrauss with binary coins," *J. Comput. Syst. Sci.*, vol. 66, no. 4, pp. 671–687, Jun 2003.

[107] P. Stoica, Z. Wang, and J. Li, "Robust Capon beamforming," *IEEE Signal Process. Lett.*, vol. 10, no. 6, pp. 172–175, Jun 2003.

[108] J. Li, P. Stoica, and Z. Wang, "On robust Capon beamforming and diagonal loading," *IEEE Trans. Signal Process.*, vol. 51, no. 7, pp. 1702–1715, Jul 2003.

[109] A. Swindlehurst and T. Kailath, "A performance analysis of subspace-based methods in the presence of model errors - part I: The MUSIC algorithm," *IEEE Trans. Signal Process.*, vol. 40, no. 7, pp. 1758–1774, Jul 1992.

[110] N. Chu, J. Picheral, and A. Mohammad-Djafari, "Bayesian sparse regularization in near-field wideband aeroacoustic imaging for wind tunnel test," in *Proc. Acoust. 2012 Nantes Conf. (ACOUSTICS)*, Société Française d'Acoustique, Ed., Nantes, France, Apr 2012, pp. 1391–1396.

[111] R. Giryes, M. Elad, and Y. C. Eldar, "The projected GSURE for automatic parameter tuning in iterative shrinkage methods," *Appl. Comput Harmonic Anal.*, vol. 30, no. 3, pp. 407–422, May 2011.

[112] X. Xu, X. Wei, and Z. Ye, "DOA estimation based on sparse signal recovery utilizing weighted $l_1$-norm penalty," *IEEE Signal Process. Lett.*, vol. 19, no. 3, pp. 155–158, Jan 2012.

[113] C. Mecklenbrauker, P. Gerstoft, A. Panahi, and M. Viberg, "Sequential Bayesian sparse signal reconstruction using array data," *IEEE Trans. Signal Process.*, vol. 61, no. 24, pp. 6344–6354, Dec 2013.

[114] J. Huang, M. Dong, and S. Li, "A new method of regularization parameter estimation for source localization," in *Proc. IEEE CIE Int. Conf. Radar (RADAR)*, Chengdu, China, Oct 2011, vol. 2, pp. 1804–1808.

[115] C. Zheng, G. Li, H. Zhang, and X. Wang, "An approach of regularization parameter estimation for sparse signal recovery," in *Proc. IEEE 10th Int. Conf. Signal Process. (ICSP)*, Beijing, China, Nov 2010, pp. 385–388.

[116] J. J. Fuchs, "Linear programming in spectral estimation. application to array processing," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Atlanta, Georgia, USA, May 1996, vol. 6, pp. 3161–3164.

[117] Z.-M. Liu and Y.-Y. Zhou, "A unified framework and sparse Bayesian perspective for direction-of-arrival estimation in the presence of array imperfections," *IEEE Trans. Signal Process.*, vol. 61, no. 15, pp. 3786–3798, Aug 2013.

[118] C. Hansen and D. P. O'Leary, "The use of the L-curve in the regularization of discrete ill-posed problems," *SIAM J. Sci. Comput.*, vol. 14, no. 6, pp. 1487–1503, Nov 1993.

[119] B. D. Rao, K. Engan, S. F. Cotter, J. Palmer, and K. Kreutz-Delgado, "Subset selection in noise based on diversity measure minimization," *IEEE Trans. Signal Process.*, vol. 51, no. 3, pp. 760–770, Mar 2003.

[120] N. Chu, J. Picheral, and A. Mohammad-Djafari, "A robust super-resolution approach with sparsity constraint for near-field wideband acoustic imaging.," in *Proc. IEEE Symp. Signal Process. Inf. Technol. (ISSPIT)*, Bilbao, Spain, Dec 2011, pp. 310–315.

[121] J. Zheng and M. Kaveh, "Directions-of-arrival estimation using a sparse spatial spectrum model with uncertainty," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Prague, Czech Republic, May 2011, pp. 2848–2851.

[122] W. Shi, J. Zheng, M. Kaveh, and J. Huang, "Robust sparse spectral fitting in element and beam spaces for directions-of-arrival and power estimation," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Kyoto, Japan, Mar 2012, pp. 2705–2708.

[123] D. D. Feldman and L. J. Griffiths, "A projection approach for robust adaptive beamforming," *IEEE Trans. Signal Process.*, vol. 42, no. 4, pp. 867–876, Apr 1994.

[124] B. Liao, S.-C. Chan, and K.-M. Tsui, "Recursive steering vector estimation and adaptive beamforming under uncertainties," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 49, no. 1, pp. 489–501, Jan 2013.

[125] J. F. Sturm, "Using sedumi 1.02, a matlab toolbox for optimization over symmetric cones," 1998.

[126] Z. Cvetkovski, *Inequalities: Theorems, Techniques and Selected Problems*, Springer, 2012.

[127] G. Grimmett and D. Stirzaker, *Probability And Random Processes*, Oxford University Press, 3rd edition, 2001.

[128] S. Nemirovsky and M. A. Doron, "Sensitivity of MUSIC and root-MUSIC to gain calibration errors of 2D arbitrary array configuration," in *Proc. 3rd Sensor Array and Multichannel Signal Processing Workshop (SAM)*, Barcelona, Spain, Jul 2004, pp. 594–598.

[129] J. W. Pierre, R. Pappu, and R. F. Kubichek, "Effects of quadrature receiver gain error on direction-finding algorithms," in *Proc. IEEE Pacific Rim Conf. Commun. Comput. Signal Process. (PaRim)*, Victoria, Canada, May 1993, vol. 2, pp. 378–381.

[130] M. S. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret, "Applications of second-order cone programming," *Linear Algebra and its Applications*, vol. 284, pp. 193 – 228, Nov 1998.

[131] C. Weiss and A. M. Zoubir, "Dictionary learning strategies for compressed fiber sensing using a probabilistic sparse model," submitted to IEEE Trans. Signal Process. in December 2016. Online available: arXiv:1610.06902 [stat.ML].

[132] M. Mishali, Y. C. Eldar, O. Dounaevsky, and E. Shoshan, "Xampling: Analog to digital at sub-Nyquist rates," *IET Circuits Devices Syst.*, vol. 5, no. 1, pp. 8–20, Jan 2011.

[133] Y. Liang, M. Chen, H. Chen, C. Lei, P. Li, and S. Xie, "Photonic-assisted multi-channel compressive sampling based on effective time delay pattern," *Opt. Express*, vol. 21, no. 22, pp. 25700–25707, Nov 2013.

[134] R. Yao, Q. Pian, and X. Intes, "Molecular fluorescence tomography with structured light and compressive sensing," in *Opt. Life Sci.* 2015, p. JT3A.25, OSA.

[135] C. Lu, J. Shi, and J. Jia, "Online robust dictionary learning," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit. (CVPR)*, Portland, Oregon, USA, Jun 2013, pp. 415–422.

[136] C. D. Austin, R. L. Moses, J. N. Ash, and E. Ertin, "On the relation between sparse reconstruction and parameter estimation with model order selection," *IEEE J. Sel. Top. Signal Process.*, vol. 4, no. 3, pp. 560–570, Jun 2010.

[137] C. D. Austin, J. N. Ash, and R. L. Moses, "Dynamic dictionary algorithms for model order and parameter estimation," *IEEE Transactions on Signal Processing*, vol. 61, no. 20, pp. 5117–5130, Oct 2013.

[138] H. Raja, W. U. Bajwa, F. Ahmad, and M. G. Amin, "Parametric dictionary learning for TWRI using distributed particle swarm optimization," in *Proc. IEEE Radar Conf. (RADAR)*, Philadelphia, Pennsylvania, USA, May 2016.

[139] Z. Ben-Haim and Y. C. Eldar, "The Cramér-Rao bound for estimating a sparse parameter vector," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3384–3389, Jun 2010.

[140] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*, Prentice Hall, 1993.

[141] J. D. Gorman and A. O. Hero, "Lower bounds for parametric estimation with constraints," *IEEE Trans. Inf. Theory*, vol. 36, no. 6, pp. 1285–1301, Nov 1990.

[142] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, 2012.

[143] K. B. Petersen and M. S. Pedersen, "The matrix cookbook," Nov 2012, Version 20121115.

[144] K. Tamura and M. Nakazawa, "Dispersion-tuned harmonically mode-locked fiber ring laser for self-synchronization to an external clock," *Opt. Lett.*, vol. 21, no. 24, pp. 1984–1986, Dec 1996.

[145] A. E. Willner and W. Shieh, "Optimal spectral and power parameters for all-optical wavelength shifting: single stage, fanout, and cascadability," *J. Lightwave Technol.*, vol. 13, no. 5, pp. 771–781, May 1995.

[146] A. Abd El Aziz, W.P. Ng, Z. Ghassemlooy, M.H. Aly, and R. Ngah, "SOA gain uniformity improvement employing a non-uniform biasing technique for ultra-high speed optical routers," in *Proc. 7th Int. Symp. Commun. Syst. Netw. and Digit. Signal Process. (CSNDSP)*, Newcastle, UK, Jul 2010, pp. 642–647.

[147] M. J. Connelly, "Wideband semiconductor optical amplifier steady-state numerical model," *IEEE J. Quantum Electron.*, vol. 37, no. 3, pp. 439–447, Mar 2001.

[148] S. J. Savory, "Digital filters for coherent optical receivers," *Opt. Express*, vol. 16, no. 2, pp. 804–817, Jan 2008.

[149] M. Seimetz, *High-Order Modulation for Optical Fiber Transmission: Phase and Quadrature Amplitude Modulation*, Springer, 2009.

[150] E. Ip, A. P. T. Lau, D. J. F. Barros, and J. M. Kahn, "Coherent detection in optical fiber systems," *Opt. Express*, vol. 16, no. 2, pp. 753–791, Jan 2008.

[151] H. Venghaus, Ed., *Wavelength Filters in Fibre Optics*, vol. 123 of *Springer Series Opt. Sci.*, Springer, 2006.

[152] A. Yariv, "Coupled-mode theory for guided-wave optics," *IEEE J. Quantum Electron.*, vol. 9, no. 9, pp. 919–933, Sep 1973.

[153] H. Kogelnik, "Theory of optical waveguides," in *Guided-Wave Optoelectronics*, T. Tamir, Ed., vol. 26 of *Springer Series in Electronics and Photonics*, pp. 7–88. Springer, 1988.

[154] S. B. Alexander, *Optical Communication Receiver Design*, SPIE, 1997.

[155] C. D. Austin, E. Ertin, J. N. Ash, and R. L. Moses, "On the relation between sparse sampling and parametric estimation," in *Proc. IEEE 13th Digit. Signal Process. Workshop and 5th IEEE Signal Process. Edu. Workshop (DSP/SPE)*, Marco Island, Florida, USA, Jan 2009, pp. 387–392.

[156] R. Miller and C. Chang, "A modified Cramér-Rao bound and its applications," *IEEE Trans. Inf. Theory*, vol. 24, no. 3, pp. 398–400, May 1978.

[157] F. Gini and R. Reggiannini, "On the use of Cramér-Rao-like bounds in the presence of random nuisance parameters," *IEEE Trans. Commun.*, vol. 48, no. 12, pp. 2120–2126, Dec 2000.

[158] M. Leigsnering, F. Ahmad, M. G. Amin, and A. M. Zoubir, "Parametric dictionary learning for sparsity-based TWRI in multipath environments," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 52, no. 2, pp. 532–547, Apr 2016.

[159] J. Huang and T. Zhang, "The benefit of group sparsity," *Ann. Statist.*, vol. 38, no. 4, pp. 1978–2004, Aug 2010.

[160] D. Xu, L. Du, and H. Liu, "Radar signal parameter estimation with sparse Bayesian representation based on zoom-dictionary," in *Proc. IEEE Int. Conf. Comput Inf. Technol. (CIT)*, Xi'an, Shaanxi, China, Sep 2014, pp. 123–128.

[161] R. Chartrand, "Exact reconstruction of sparse signals via nonconvex minimization," *IEEE Signal Process. Lett.*, vol. 14, no. 10, pp. 707–710, Oct 2007.

[162] S. Mohamed, K. A. Heller, and Z. Ghahramani, "Bayesian and $l_1$ approaches for sparse unsupervised learning," in *Proc. 29th Int. Conf. Mach. Learn. (ICML)*, John Langford and Joelle Pineau, Eds., New York, USA, 2012, pp. 751–758, ACM.

[163] C. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.

[164] A. Yuan, "Bayesian Frequentist hybrid inference," *Ann. Stat.*, vol. 37, no. 5A, pp. 2458–2501, Oct 2009.

[165] A. Yuan, X. Zhang, and G. Han, "Review of hybrid Bayesian inference and its applications," *Ann. Biometrics Biostat.*, vol. 2, no. 3, pp. 1023, Sep 2015.

[166] H. Ishwaran and J. S. Rao, "Spike and slab variable selection: Frequentist and Bayesian strategies," *Ann. Statist.*, vol. 33, no. 2, pp. 730–773, Apr 2005.

[167] N. G. Polson and J. G. Scott, "Shrink globally, act locally: Sparse Bayesian regularization and prediction," in *Bayesian Statistics*. 2010, vol. 9, Oxford University Press.

[168] Y. C. Eldar, P. Kuppinger, and H. Bolcskei, "Block-sparse signals: Uncertainty relations and efficient recovery," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3042–3054, Jun 2010.

[169] Z. Zhang and B. D. Rao, "Sparse signal recovery with temporally correlated source vectors using sparse Bayesian learning," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 5, pp. 912–926, Sep 2011.

[170] O'Hagan, "Bayesian inference," in *Kendall's Advanced Theory of Statistics*. 1994, vol. 2B, Wiley-Blackwell.

[171] M. W. Seeger, "Bayesian inference and optimal design for the sparse linear model," *J. Mach. Learn. Res.*, vol. 9, pp. 759–813, Jun 2008.

[172] A. Mohammad-Djafari, "Bayesian approach with prior models which enforce sparsity in signal and image processing," *EURASIP J. Adv. Signal Process.*, vol. 2012, no. 1, pp. 1–19, Mar 2012.

[173] M. Carlin, P. Rocca, G. Oliveri, F. Viani, and A. Massa, "Directions-of-arrival estimation through Bayesian compressive sensing strategies," *IEEE Trans. Antennas Propag.*, vol. 61, no. 7, pp. 3828–3838, Jul 2013.

[174] S. Ji and L. Carin, "Bayesian compressive sensing and projection optimization," in *Proc. 24th Int. Conf. Machine Learn. (ICML)*, Corvallis, Oregon, USA, Jun 2007, pp. 377–384.

[175] S. D. Babacan, R. Molina, and A. K. Katsaggelos, "Bayesian compressive sensing using Laplace priors," *IEEE Trans. Image Process.*, vol. 19, no. 1, pp. 53–63, Jan 2010.

[176] T.J. Mitchell and J.J. Beauchamp, "Bayesian variable selection in linear regression," *J. Amer. Statist. Assoc.*, vol. 83, no. 404, pp. 1023–1032, 1988.

[177] H. Ishwaran and S. Rao, "Spike and slab gene selection for multigroup microarray data," *J. Amer. Statist. Assoc.*, vol. 100, pp. 764–780, Sep 2005.

[178] T. L. Hansen, P. B. Jorgensen, N. L. Pedersen, C. N. Manchón, and B. H. Fleury, "Bayesian compressed sensing with unknown measurement noise level," in *Proc. Asilomar Conf. Signals Syst. Comput. (ASILOMAR)*, Nov 2013, pp. 148–152.

[179] C. M. Carvalho, N. G. Polson, and J. G. Scott, "The horseshoe estimator for sparse signals," *Biometrika*, vol. 97, no. 2, pp. 465–480, Apr 2010.

[180] T. Peltola, A. S. Havulinna, V. Salomaa, and A. Vehtari, "Hierarchical Bayesian survival analysis and projective covariate selection in cardiovascular event risk prediction," in *Proc. 11th UAI Bayesian Modeling Applicat. Workshop, CEUR Workshop Proc.*, Quebec, Canada, Jul 2014, vol. 1218, pp. 79–88.

[181] M. D. Gupta and S. Kumar, "Non-convex p-norm projection for robust sparsity," in *Proc. IEEE Int. Conf. Comput. Vision (ICCV)*, Sydney, Australia, 2013, pp. 1593–1600.

[182] R. Chartrand and V. Staneva, "Restricted isometry properties and nonconvex compressive sensing," *Inverse Problems*, vol. 24, no. 3, pp. 035020, Jun 2008.

[183] M. B. Wakin, S. Sarvotham, M. F. Duarte, D. Baron, and R. G. Baraniuk, "Recovery of jointly sparse signals from few random projections," in *Proc. Workshop Neural Inform. Process. Syst. (NIPS)*, Vancouver, Canada, Dec 2005, pp. 1435–1442.

[184] J. Chen, F. D. Bushman, J. D. Lewis, G. D. Wu, and H. Li, "Structure-constrained sparse canonical correlation analysis with an application to microbiome data analysis," *Biostatistics*, vol. 14, pp. 244–258, Apr 2013.

[185] Y. Altmann, M. Pereyra, and J. Bioucas-Dias, "Collaborative sparse regression using spatially correlated supports - application to hyperspectral unmixing," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5800–5811, Dec 2015.

[186] C. Papageorgiou, F. Girosi, and T. Poggio, "Sparse correlation kernel reconstruction," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Phoenix, Arizona, USA, Mar 1999, vol. 3, pp. 1633–1636.

[187] P. Misra, W. Hu, M. Yang, and S. Jha, "Efficient cross-correlation via sparse representation in sensor networks," in *Proc. ACM/IEEE 11th Int. Conf. Inform. Process. Sensor Networks (IPSN)*, Beijing, China, Apr 2012, pp. 13–24.

[188] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*, MIT Press, 2012.

[189] R. M. Neal, *MCMC Using Hamiltonian Dynamics*, Handbook of Markov Chain Monte Carlo. Chapman and Hall / CRC, 2011, Chapter 5, S. Brooks and A. Gelman and G. Jones and X.-L. Meng, Eds.

[190] D. Fink, "A compendium of conjugate priors. In progress report: Extension and enhancement of methods for setting data quality objectives. (DOE contract 95-831)," Tech. Rep., 1995.

[191] M. D. Homan and A. Gelman, "The No-U-turn Sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1593–1623, Jan 2014.

[192] Stan Development Team, "Stan: A C++ library for probability and sampling, version 2.9.0," 2015.

# Curriculum Vitae

| | |
|---|---|
| Name: | Christian Weiß |
| Date of birth: | 21st June, 1985 |
| Place of birth: | Wiesbaden, Germany |
| Family status: | single |

### Education

| | |
|---|---|
| since 06/2012 | Technische Universität Darmstadt, Germany<br>Graduate School of Computational Engineering<br>Affiliate Research Member |
| 08/2015 - 10/2015 | National Tsing Hua University, Taiwan, R.O.C.<br>Wireless Communications and Signal Processing Lab.<br>Visiting Researcher |
| 10/2009 - 11/2011 | Technische Universität Darmstadt, Germany<br>Electrical Engineering and Information Technology<br>Master of Science (M.Sc.) |
| 10/2010 - 03/2011 | The University of Tokyo, Japan<br>Independent Research:<br>"Digital Signal Processing for Optical Communications and Transmission Systems" |
| 10/2005 - 09/2009 | Technische Universität Darmstadt, Germany<br>Information and Communications Technology<br>Bachelor of Science (B.Sc.) |
| 06/2004 | High school degree (Abitur) at<br>Humbold Gymnasium, Wiesbaden, Germany |

### Work Experience

| | |
|---|---|
| since 01/2012 | Research Associate at<br>Signal Processing Group<br>Institute of Telecommunications<br>Technische Universität Darmstadt, Germany |
| 01/2009 - 12/2010 | Software Developer at<br>iTecConsult Kohrt, Munich, Germany |
| 10/2008 - 03/2009 | Industrial Trainee at<br>Rohde & Schwarz GmbH, Munich, Germany |

Erklärung laut §9 der Promotionsordnung

Ich versichere hiermit, dass ich die vorliegende Dissertation allein und nur unter Verwendung der angegebenen Literatur verfasst habe. Die Arbeit hat bisher noch nicht zu Prüfungszwecken gedient.

Darmstadt, 06. Dezember 2016,